(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) Publication number: **0 499 003 A1**

(12) EUROPEAN PATENT APPLICATION

(21) Application number: **91400388.4**

(22) Date of filing: **14.02.91**

(51) Int. Cl.[5]: **C12N 15/31**, C12Q 1/68, G01N 33/569, A61K 39/04, A61K 39/395, C12N 15/62, C12P 21/02, C12N 1/21

(43) Date of publication of application:
19.08.92 Bulletin 92/34

(84) Designated Contracting States:
GB

(71) Applicant: **N.V. INNOGENETICS S.A.**
Industriepark Zwijnaarde 7, Box 4
B-9710 Gent(BE)

(72) Inventor: **Content, Jean**
5 avenue Simonne
B-1640 Rhode St-Genèse(BE)

Inventor: **De Wit, Lucas**
Victor Vergauwenstraat 46
B-2870 Puurs(BE)
Inventor: **De Bruyn, Jacqueline**
192 rue de Hongrie
1650 Beersel(BE)

(74) Representative: **Grosset-Fournier, Chantal Catherine et al**
ERNEST GUTMANN-YVES PLASSERAUD S.A.,
67 boulevard Haussmann
F-75008 Paris(FR)

(54) Polypeptides and peptides, particularly recombinant polypeptides and peptides, nucleic acids coding for the same and use of these polypeptides and peptides in the diagnosis of tuberculosis.

(57) The invention relates :
- to nucleic acids which contain particularly a nucleotide sequence extending from the extremity constituted by the nucleotide at position (1) to the extremity constituted by the nucleotide at position (149) represented on Figure 1,
- to the polypeptides coded by said nucleic acids.

The polypeptides of the invention can be used for the diagnosis of tuberculosis, and can also be part of the active principle in the preparation of vaccine against tuberculosis.

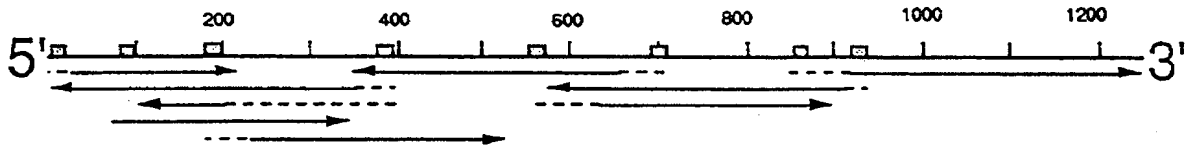


Figure 1

EP 0 499 003 A1

The invention relates to polypeptides and peptides, particularly recombinant polypeptides and peptides, which can be used for the diagnosis of tuberculosis. The invention also relates to a process for preparing the above-said polypeptides and peptides, which are in a state of biological purity such that they can be used as part of the active principle in the preparation of vaccines against tuberculosis.

It also relates to nucleic acids coding for said polypeptides and peptides.

Furthermore, the invention relates to the in vitro diagnostic methods and kits using the above-said polypeptides and peptides and to the vaccines containing the above-said polypeptides and peptides as active principle against tuberculosis.

By "recombinant polypeptides or peptides" it is to be understood that it relates to any molecule having a polypeptidic chain liable to be produced by genetic engineering, through transcription and translation, of a corresponding DNA sequence under the control of appropriate regulation elements within an efficient cellular host. Consequently, the expression "recombinant polypeptides" such as is used herein does not exclude the possibility for the polypeptides to comprise other groups, such as glycosylated groups.

The term "recombinant" indeed involves the fact that the polypeptide has been produced by genetic engineering, particularly because it results from the expression in a cellular host of the corresponding nucleic acid sequences which have previously been introduced into the expression vector used in said host.

Nevertheless, it must be understood that this expression does not exclude the possibility for the polypeptide to be produced by a different process, for instance by classical chemical synthesis according to methods used in the protein synthesis or by proteolytic cleavage of larger molecules.

The expression "biologically pure" or "biological purity" means on the one hand a grade of purity such that the recombinant polypeptide can be used for the production of vaccinating compositions and on the other hand the absence of contaminants, more particularly of natural contaminants.

Tuberculosis remains a major disease in developing countries. The situation is dramatic in some countries, particularly where high incidence of tuberculosis among AIDS patients represents a new source of dissemination of the disease.

Tuberculosis is a chronic infectious disease in which cell-mediated immune mechanisms play an essential role both for protection against and control of the disease.

Despite BCG vaccination, and some effective drugs, tuberculosis remains a major global problem. Skin testing with tuberculin PPD (protein-purified derivative) largely used for screening of the disease is poorly specific, due to cross reactivity with other pathogenic or environmental saprophytic mycobacteria.

Moreover, tuberculin PPD when used in serological tests (ELISA) does not allow to discriminate between patients who have been vaccinated by BCG, or those who have been primo-infected, from those who are developing evolutive tuberculosis and for whom an early and rapid diagnosis would be necessary.

A protein with a molecular weight of 32-kDa has already been purified from zinc deficient M. bovis BCG culture filtrate. This protein was identified as antigen 85A (De Bruyn J. et al., 1987, "Purification, partial characterization and identification of a 32-kDa protein antigen of Mycobacterium bovis BCG" Microb. Pathogen. 2:351). Its NH_2-terminal amino acid sequence (Phe-Ser-Arg-Pro-Gly-Leu) is identical to that reported for the α-antigen (antigen 85B) protein purified from M. Bovis BCG substrain Tokyo (Wiker, H.G. et al., 1986, "MPB59, a widely cross-reacting protein of Mycobacterium bovis BCG" Int. Arch. Allergy Appl. Immunol. 81:307) . The antigen 85-complex is present among different strains of mycobacteria (De Bruyn J. et al., 1989, "Effect of zinc deficiency of the appearance of two immunodominant protein antigens (32-kDa and 65-kDa) in culture filtrates of Mycobacteria" J. Gen Microbiol. 135:79). It is secreted by living bacilli as a predominant protein in normal Sauton culture filtrate and could be useful in the serodiagnosis of tuberculosis (Turneer M. et al., 1988, "Humoral immune response in human tuberculosis: immunoglobulins G, A and M directed against the purified P32 protein antigen of Mycobacterium bovis bacillus Calmette-Guérin" J. Clin. Microbiol. 26:1714) and leprosy (Rumschlag H.S. et al., 1988, "Serological response of patients with lepromatous and tuberculosis leprosy to 30-, 31- and 32-kilodalton antigens of Mycobacterium tuberculosis" J. Clin. Microbiol. 26:2200). Furthermore, the 32-kDa protein induces specific lymphoproliferation and interferon-γ(IFN-γ) production in peripheral blood leucocytes from tuberculosis (Huygen K. et al., 1988, "Specific lymphoproliferation, γ-interferon production and serum immunoglobulin G directed against a purified 32-kDa mycobacterial antigen (P32) in patients with active tuberculosis" Scand. J. Immunol. 27:187), and leprosy patients and from PPD- and lepromin-positive healthy subjects. Recent findings indicate that the amount of 32 kDa protein induced IFN-γ in BCG-sensitized mouse spleen cells is under probable H-2 control (Huygen K. et al, 1989, "H-2-linked control of in vitro γ interferon production in response to a 32-kilodalton antigen (P32) of Mycobacterium bovis bacillus Calmette-Guérin" Infect. Imm. 56:3196). Finally, the high affinity of mycobacteria for fibronectin is related to proteins of the antigen 85-complex (Abou-Zeid C. et al., 1988, "Characterization of fibronectin-binding antigens released by Mycobacterium tuberculosis and Mycobacterium bovis BCG" Infect. Imm. 56:3046).

Wiker et al. (Wiker H.G. et al., 1990, "Evidence for three separate genes encoding the proteins of the mycobacterial antigen 85 complex" Infect. Immun. 58:272) showed recently that the antigens 85A, B and C isolated from M. bovis BCG culture filtrate present a few amino acid replacements in their NH_2 terminal region strongly suggesting the existence of multiple genes coding for these proteins. But, the data given for the antigen 85C of M. bovis BCG are insufficient to enable its unambiguous identifiability as well as the characterization of its structural and functional elements.

The gene encoding the α-antigen from Mycobacterium bovis BCG has been described (Borremans L. et al., 1989, "Cloning, sequence determination and expression of a 32-Kilodalton protein gene of Mycobacterium tuberculosis" Infect. Immun. 57:3123) which presented 77.5% homology at the DNA level within the coding region with the α-antigen gene (Matsuo K. et al., 1988, "Cloning and expression of the Mycobacterium bovis BCG gene for extracellular α-antigen" J. Bacteriol. 170:3847). Moreover, we have recently isolated and sequenced a corresponding 32-kDa protein genomic clone from our λgt11 BCG library (prepared from strain M. bovis BCG 1173P2). The complete sequence of this gene is identical with that from Mycobacterium tuberculosis except for a single silent nucleotide change (De Wit L. et al., 1990, "Nucleotide sequence of the 32 kDa-protein gene (antigen 85A) of Mycobacterium bovis BCG" Nucl. Ac. Res. 18:3995). This confirmed the previous findings that partial protein sequence of several tryptic peptides derived from highly purified 32-kDa protein from M. bovis BCG present the characteristic 85A sequence (Borremans L. et al., 1989, "Cloning, sequence determination and expression of a 32-Kilodalton protein gene of Mycobacterium tuberculosis" Infect. Immun. 57:3123 ; De Wit L. et al., 1990, "Nucleotide sequence of the 32 kDa-protein gene (antigen 85A) of Mycobacterium bovis BCG" Nucl. Ac. Res. 18:3995) and not the 85B sequence. Thus so far it was likely, but not demonstrated, that the genome of M. bovis BCG contained at least two genes coding for antigen 85A and 85B respectively. As to the genome of the Mycobacterium tuberculosis, nothing was proved as to the existence of new genes, besides the genes coding respectively for 85A and 85B.

An aspect of the invention is to provide with a new family of nucleic acids coding for new proteins and polypeptides which can be used for the detection and control of tuberculosis.

Another aspect of the invention is to provide nucleic acids coding for the peptidic chains of biologically pure recombinant polypeptides which enable their preparation on a large scale.

Another aspect of the invention is to provide antigens which can be used in serological tests as an in vitro rapid diagnostic test for tuberculosis.

Another aspect of the invention is to provide a rapid in vitro diagnostic means for tuberculosis, enabling it to discriminate between patients suffering from an evolutive tuberculosis from those who have been vaccinated against BCG or who have been primo-infected.

Another aspect of the invention is to provide nucleic probes which can be used as in vitro diagnostic reagent for tuberculosis as well as in vitro diagnostic reagent for identifying M. tuberculosis from other strains of mycobacteria.

The nucleic acids of the invention

* contain a nucleotide sequence extending from the extremity constituted by the nucleotide at position (1) to the extremity constituted by the nucleotide at position (149) represented on Figure 1,
* or contain one at least of the nucleotide sequences coding for the following peptides or polypeptides :
  - the one extending from the extremity constituted by amino acid at position (-46) to the extremity constituted by amino acid at position (-1) represented on Figure 1, or
  - the one extending from the extremity constituted by amino acid at position (-21) to the extremity constituted by amino acid at position (-1) represented on Figure 1, or
  - SQSNGQNY, or
  - PMVQIPRLVA, or
  - GSTLRTNQTPRDTYAADGGRNG or
  - PPAAPAAPAA,
* or contain nucleotidic sequences :
  - hybridizing with the above-mentioned nucleotide sequences, or their complements,
  - complementary to the above-mentioned nucleotide sequences, or
  - which are the above-mentioned nucleotide sequences wherein T can be replaced by U,
* or are constituted by the above-mentioned nucleotide sequences.

The hybridization takes place under the following conditions :

- hybridization and wash medium :
  * a preferred hybridization medium contains about 3 x SSC [SSC = 0.15 M sodium chloride, 0.015 M sodium citrate, pH 7] about 25 mM of phosphate buffer pH 7.1, and 20% deionized formamide, 0.02% Ficoll, 0.02% BSA, 0.02% polyvinylpyrrolidone and about 0.1 mg/ml sheared denatured

salmon sperm DNA,

* a preferred wash medium contains about 3 x SSC, about 25 mM phosphate buffer, pH 7.1 and 20% deionized formamide;

- hybridization temperature (HT) and wash temperature (WT) are between 45 ° C and 65 ° C;
- for the nucleotide sequence extending from the extremity constituted by the nucleotide at position (1) to the extremity constituted by the nucleotide at position (149) represented on Figure 1 :
  HT = WT = 65 ° C
  for the nucleic acids of the invention defined by coded polypeptides X - Y : i.e.
  . the sequence extending from the extremity constituted by the amino acid at position (X) to the extremity constituted by the amino acid at position (Y) represented on Figure 1,
  . the sequence extending from the extremity constituted by the amino acid at position (-46) to the extremity constituted by the amino acid at position (-1) represented on Figure 1,
  HT = WT = 65 ° C
  . the sequence extending from the extremity constituted by the amino acid at position (-21) to the extremity constituted by the amino acid at position (-1) represented on Figure 1,
  HT = WT = 60 ° C
  for the nucleic acids defined by coded polypeptides represented by their sequence :
  . SQSNGQNY HT = WT = 45 ° C
  . PMVQIPRLVA HT = WT = 55 ° C
  . GLTLRTNQTPRDTYAAGGRNG HT = WT = 65 ° C
  . PPAAPAAPAA HT = WT = 65 ° C.

The above-mentioned temperatures are to be expressed as approximately ± 5 ° C.

Advantageous nucleic acids of the invention contain one at least of the following nucleotide sequences :

- the one extending from the extremity constituted by the nucleotide at position (150) to the extremity constituted by the nucleotide at position (287) on Figure 1,
- the one extending from the extremity constituted by the nucleotide at position (224) to the extremity constituted by the nucleotide at position (287) on Figure 1,
- the one extending from the extremity constituted by the nucleotide at position (537) to the extremity constituted by the nucleotide at position (560) on Figure 1,
- the one extending from the extremity constituted by the nucleotide at position (858) to the extremity constituted by the nucleotide at position (887) on Figure 1,
- the one extending from the extremity constituted by the nucleotide at position (972) to the extremity constituted by the nucleotide at position (1037) on Figure 1,
- the one extending from the extremity constituted by the nucleotide at position (1140) to the extremity constituted by the nucleotide at position (1169) on Figure 1, or contain nucleotidic sequences :
- hybridizing with the above-mentioned nucleotide sequences, or
- complementary to the above-mentioned nucleotide sequences, or
- which are the above-mentioned nucleotide sequences wherein T can be replaced by U, or are constituted by the above-mentioned nucleotide sequences.

The hybridization takes place under the following conditions :

- hybridization and wash medium are as above defined;
- hybridization temperature (HT) and wash temperature (WT) for the nucleic acids of the invention defined by X - Y : i.e. by the sequence extending from the extremity constituted by the nucleotide at position (X) to the extremity constituted by the nucleotide at position (Y) represented on Figure 1 :
  (150) - (287) HT = WT = 65 ° C
  (224) - (287) HT = WT = 60 ° C
  (537) - (560) HT = WT = 45 ° C
  (858) - (887) HT = WT = 55 ° C
  (972) - (1037) HT = WT = 65 ° C
  (1140) - (1169) HT = WT = 65 ° C.

An advantageous group of nucleic acids of the invention contains the nucleotide sequence coding for the following peptide :
SQSNGQNY
and liable to hybridize with the following nucleotide sequence :
CGGCTGGGAC(or T)ATCAACACCCCGGC
and liable to hybridize neither with
GCCTGCGGCAAGGCCGGTTGCCAG
nor with

GCCTGCGGTAAGGCTGGCTGCCAG
nor with
GCCTGCGGCAAGGCCGGCTGCACG
or are constituted by the above-mentioned nucleotide sequences.

The above-mentioned hybridization can take place when the hybridization and wash medium is as indicated above; and the hybridization and wash temperature is 52 ° C.

The expression "not liable to hybridize with" means that the nucleic acid molecule of the invention does not contain a stretch of nucleotide hybridizing at 52 ° C in the above defined medium with the three probes defined above.

Advantageous nucleic acids of the invention contain one at least of the above-mentioned nucleotide sequences or are constituted by the above-mentioned nucleotide sequences and besides contain an open reading frame coding for a polypeptide

- liable to react selectively with human sera from tuberculosis patients and particularly patients developing an evolutive tuberculosis,
- or liable to be recognized by antibodies also recognizing the amino acid sequence extending from the extremity constituted by amino acid at position (1) to the extremity constituted by amino acid at position (294) represented on Figure 1,
- or liable to generate antibodies recognizing the amino acid sequence extending from the extremity constituted by amino acid at position (1) to the extremity constituted by amino acid at position (294) represented on Figure 1.

The recognition of the above-mentioned sequence of the 294 amino acids (or of the polypeptides of the invention) by the abovesaid antibodies means that the abovesaid sequence forms a complex with one of the above-mentioned antibodies.

Forming a complex between the antigen (i.e. the sequence of 294 amino acids or any polypeptide of the invention) and the antibodies and detecting the existence of a formed complex can be done according to classical techniques (such as the one using a tracer labeled with radioactive isotopes or an enzyme).

Hereafter is given, in a non limitative way, a process for testing the selective reaction between the antigen and human sera from tuberculosis patients and particularly patients developing an evolutive tuberculosis.

This test is an immunoblotting (Western blotting) analysis, in the case where the polypeptides of the invention are obtained by recombinant techniques. This test can also be used for polypeptides of the invention obtained by a different preparation process. After sodium dodecyl sulfate-polyacrylamide gel electrophoresis, polypeptides of the invention are blotted onto nitrocellulose membranes (Hybond c. (Amersham)) as described by Towbin H. et al., 1979, "Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications" Proc. Natl. Acad. Sci. USA 76:4350-4354. The expression of polypeptides of the invention fused to β-galactosidase in E. coli Y1089, is visualized by the binding of a polyclonal rabbit anti-32-kDa BCG protein serum (1:1,000) or by using a monoclonal anti-β-galactosidase antibody (Promega). The secondary antibody (alkaline phosphatase anti-rabbit immunoglobulin G and anti-mouse alkaline phosphatase immunoglobulin G conjugates, respectively) is diluted as recommended by the supplier (Promega).

In order to identify selective recognition of polypeptides of the invention and of fusion proteins of the invention by human tuberculous sera, nitrocellulose sheets are incubated overnight with these sera (1:50) (after blocking a specific protein-binding sites). The human tuberculous sera are selected for their reactivity (high or low) against the purified 32-kDa antigen of BCG tested in a dot blot assay as described in Van Vooren J.P. et al., 1989, "Local anti-P32 humoral response in tuberculous meningitis". Tubercle. 70:123-126. Reactive areas on the nitrocellulose sheets are revealed by incubation with peroxidase conjugated goat anti-human immunoglobulin G antibody (Dakopatts, Copenhagen, Denmark) (1:200) for 4h, and after repeated washings, color reaction is developed by adding peroxidase substrate (α-chloronaphtol)(Bio-Rad Laboratories, Richmond, Calif.) in the presence of peroxidase and hydrogen peroxide.

Advantageous nucleic acids of the invention contain or are constituted by one of the above-mentioned nucleotide sequences, contain an open reading frame and code for a mature polypeptide of about 30 to about 35 kD, and contain a sequence coding for a signal sequence.

Advantageous nucleic acids of the invention contain one at least of the nucleotide sequences coding for the following polypeptides :

- the one extending from the extremity constituted by amino acid at position (-46) to the extremity constituted by amino acid at position (-1) represented on Figure 1, or
- the one extending from the extremity constituted by amino acid at position (-21) to the extremity constituted by amino acid at position (-1) represented on Figure 1, or

- the one extending from the extremity constituted by amino acid at position (-46) to the extremity constituted by amino acid at position (294) represented on Figure 1, or
- the one extending from the extremity constituted by amino acid at position (-21) to the extremity constituted by amino acid at position (294) represented on Figure 1, or
- the one extending from the extremity constituted by amino acid at position (1) to the extremity constituted by amino acid at position (294) represented on Figure 1,

or contain nucleotidic sequences :

- hybridizing with the above-mentioned nucleotide sequences, or
- complementary to the above-mentioned nucleotide sequences, or
- which are the above-mentioned nucleotide sequences wherein T can be replaced by U, or are constituted by the above-mentioned nucleotide sequences.

The hybridization takes place under the following conditions :

- hybridization and wash medium are as above defined;
- hybridization temperature (HT) and wash temperature (WT) for the nucleic acids of the invention defined by coded polypeptides X - Y: i.e. by the coded sequence extending from the extremity constituted by the amino acid at position (X) to the extremity constituted by the amino acid at position (Y) represented on Figure 1:

(-46) - (-1) HT = WT = 65 °C
(-21) - (-1) HT = WT = 60 °C
(-46) - (294) HT = WT = 70 °C
(-21) - (294) HT = WT = 70 °C
(1) - (294) HT = WT = 70 °C.

Advantageous nucleic acids of the invention contain one at least of the following nucleotide sequences :

- the one extending from the extremity constituted by the nucleotide at position (150) to the extremity constituted by the nucleotide at position (287) represented on Figure 1, or
- the one extending from the extremity constituted by the nucleotide at position (224) to the extremity constituted by the nucleotide at position (287) represented on Figure 1, or
- the one extending from the extremity constituted by the nucleotide at position (1) to the extremity constituted by the nucleotide at position (1169) represented on Figure 1, or
- the one extending from the extremity constituted by the nucleotide at position (150) to the extremity constituted by the nucleotide at position (1169) represented on Figure 1, or
- the one extending from the extremity constituted by the nucleotide at position (224) to the extremity constituted by the nucleotide at position (1169) represented on Figure 1, or
- the one extending from the extremity constituted by the nucleotide at position (288) to the extremity constituted by the nucleotide at position (1169) represented on Figure 1,
- the one extending from the extremity constituted by the nucleotide at position (1) to the extremity constituted by the nucleotide at position (1211) represented on Figure 1,
- the one extending from the extremity constituted by the nucleotide at position (150) to the extremity constituted by the nucleotide at position (1211) represented on Figure 1,
- the one extending from the extremity constituted by the nucleotide at position (224) to the extremity constituted by the nucleotide at position (1211) represented on Figure 1,
- the one extending from the extremity constituted by the nucleotide at position (288) to the extremity constituted by the nucleotide at position (1211) represented on Figure 1,

or contain nucleotidic sequences :

- hybridizing with the above-mentioned nucleotide sequences, or
- complementary to the above-mentioned nucleotide sequences, or
- which are the above-mentioned nucleotide sequences wherein T can be replaced by U, or are constituted by one at least of the following nucleotide sequences.

The hybridization takes place under the following conditions :

- hybridization and wash medium are as above defined;
- hybridization temperature (HT) and wash temperature (WT) for the nucleic acids of the invention defined for the nucleic acids of the invention defined by X - Y: i.e. by the sequence extending from the extremity constituted by the nucleotide at position (X) to the extremity constituted by the nucleotide at position (Y) represented on Figure 1 :

(150) - (287) HT = WT = 65 °C
(224) - (287) HT = WT = 60 °C
(150) - (1169) HT = WT = 70 °C
(1) - (1169) HT = WT = 70 °C

(224) - (1169) HT = WT = 70 °C
(288) - (1169) HT = WT = 70 °C

An advantageous nucleic acid sequence of the invention contain a nucleotide sequence coding for a polypeptide sequence extending from the extremity constituted by the amino acid at position (18) to the extremity constituted by the amino acid at position (99) represented on Figure 2B, on the fifth line, or is constituted by this nucleotide sequence.

An advantageous nucleic acid sequence of the invention contain a nucleotide sequence constituted by the nucleotide at position (584) to the extremity constituted by the nucleotide at position (828) represented on Figure 2A, on the fifth line, or is constituted by this nucleotide sequence.

The hybridization takes place under the following conditions :

- hybridization and wash medium are as above defined,
  HT = WT = 65 °C.

The invention relates also to the polypeptides coded by the nucleic acids of the invention above defined.

Advantageous polypeptides of the invention contain in their polypeptide chain, one at least of the following amino acid sequences :

- the one extending from the extremity constituted by amino acid at position (-46) to the extremity constituted by amino acid at position (-1) represented on Figure 1,
- or the one extending from the extremity constituted by amino acid at position (-21) to the extremity constituted by amino acid at position (-1) represented on Figure 1,
- SQSNGQNY,
- PMVQIPRLVA,
- GLTLRTNQTPRDTYAADGGRNG,
- PPAAPAAPAA,

or are constituted by the above-mentioned polypeptide sequences.

Advantageous polypeptides of the invention contain in their polypeptide chain, one at least of the following amino acid sequence :
SQSNGQNY
and the amino acid sequence
GWDINTPA
and containing not the amino acid sequence
ACGKAGCQ
and not the amino acid sequence
ACGKAGCT.

Interesting polypeptides are :

SQSNGQNY

GWDINTPA.

Advantageous polypeptides of the invention are liable to react selectively with human sera from tuberculosis patients and particularly patients developing an evolutive tuberculosis,
or liable to be recognized by antibodies also recognizing the polypeptide sequence extending from the extremity constituted by amino acid at position (1) to the extremity constituted by amino acid at position (294) represented on Figure 1,
or liable to generate antibodies recognizing the polypeptidic sequence extending from the extremity constituted by amino acid at position (1) to the extremity constituted by amino acid at position (294) represented on Figure 1.

The invention also includes the peptidic sequences resulting from the modification by substitution and/or by addition and/or by deletion of one or several amino acids in the above defined polypeptides and peptides in so far as this modification does not alter the following properties :
selective reaction with human sera from tuberculosis patients and particularly patients developing an evolutive tuberculosis,
and/or reaction with antibodies raised against the amino acid sequence extending from the extremity constituted by amino acid at position (1), to the extremity constituted by amino acid at position (294) represented on Fig. 1.

Advantageous polypeptides of the invention contain or are constituted by one of the above-mentioned polypeptide sequences, and are about 30 to about 35 kD and are preceded by a signal peptide.

Advantageous polypeptides of the invention contain in their polypeptide chain, one at least of the following amino acid sequences or are constituted by one of the following amino acid sequences :

- the one extending from the extremity constituted by amino acid at position (1) to the extremity constituted by amino acid at position (294) represented on Figure 1,
- the one extending from the extremity constituted by amino acid at position (-46) to the extremity constituted by amino acid at position (294) represented on Figure 1,
- the one extending from the extremity constituted by amino acid at position (-21) to the extremity constituted by amino acid at position (294) represented on Figure 1,
- the one extending from the extremity constituted by amino acid at position (-46) to the extremity constituted by amino acid at position (-1) represented on Figure 1,
- the one extending from the extremity constituted by amino acid at position (-21) to the extremity constituted by amino acid at position (-1) represented on Figure 1.

Advantageous polypeptides of the invention contain in their polypeptide chain the amino acid sequence extending from the extremity constituted by amino acid at position (18) to the extremity constituted by amino acid at position (99) represented on Figure 2B, on the fifth line.

It goes without saying that the free reactive functions which are present in some of the amino acids, which are part of the constitution of the polypeptides of the invention, particularly the free carboxyl groups which are carried by the groups Glu or by the C-terminal amino acid on the one hand and/or the free NH_2 groups carried by the N-terminal amino acid or by amino acid inside the peptidic chain, for instance Lys, on the other hand, can be modified in so far as this modification does not alter the above mentioned properties of the polypeptide.

The molecules which are thus modified are naturally part of the invention. The above mentioned carboxyl groups can be acylated or esterified.

Other modifications are also part of the invention. Particularly, the amine or ester functions or both of terminal amino acids can be themselves involved in the bond with other amino acids. For instance, the N-terminal amino acid can be linked to a sequence comprising from 1 to several amino acids corresponding to a part of the C-terminal region of another peptide.

The polypeptides according to the invention can be glycosylated or not, particularly in some of their glycosylation sites of the type Asn-X-Ser or Asn-X-Thr, X representing any amino acid.

Other advantageous polypeptides of the invention consist in one of the following amino acid sequences :

- the one extending from the extremity constituted by amino acid at position (-46) to the extremity constituted by amino acid at position (-1) represented on Figure 1,
- or the one extending from the extremity constituted by amino acid at position (-21) to the extremity constituted by amino acid at position (-1) represented on Figure 1.

These polypeptides can be used as signal peptides, the role of which is to initiate the translocation of a protein from its site of synthesis, but which is excised during translocation.

Advantageous polypeptides of the invention are the ones constituted by :

- SQSNGQNY,
- PMVQIPRLVA,
- GLTLRTNQTPRDTYAADGGRNG,
- PPAAPAAPAA,
- the one extending from the extremity constituted by amino acid at position (1) to the extremity constituted by amino acid at position (294) represented on Figure 1,
- the one extending from the extremity constituted by amino acid at position (-46) to the extremity constituted by amino acid at position (294) represented on Figure 1,
- the one extending from the extremity constituted by amino acid at position (-21) to the extremity constituted by amino acid at position (294) represented on Figure 1,
- the one extending from the extremity constituted by amino acid at position (-46) to the extremity constituted by amino acid at position (-1) represented on Figure 1,
- the one extending from the extremity constituted by amino acid at position (-21) to the extremity constituted by amino acid at position (-1) represented on Figure 1,
- GWDINTPA,
- the one extending from the extremity constituted by amino acid at position (18) to the extremity constituted by amino acid at position (99) represented on Figure 2B, on the fifth line.

It is to be noted that the above mentioned polypeptides are derived from the expression products of a DNA derived, as explained hereafter in the examples,

- from the nucleotide sequence coding for a protein of 33-kDa secreted by Mycobacterium tuberculosis

or

- from the nucleotide sequence coding for a protein secreted by M. bovis BCG, or
- from related nucleotide sequences which will be hereafter designated by 85-C genes.

The invention also relates to the amino acid sequences constituted by the above mentioned polypeptides and a protein or an heterologous sequence with respect to said polypeptide, said protein or heterologous sequence comprising for instance from about 1 to about 1000 amino acids. These amino acid sequences will be called fusion proteins.

In an advantageous fusion protein of the invention, the heterologous protein is β-galactosidase.

The invention also relates to any recombinant nucleic acids containing at least a nucleic acid of the invention inserted in an heterologous nucleic acid.

The invention relates more particularly to recombinant nucleic acid such as defined, in which the nucleotide sequence of the invention is preceded by a promoter (particularly an inducible promoter) under the control of which the transcription of said sequence is liable to be processed and possibly followed by a sequence coding for transcription termination signals.

The invention also relates to the recombinant nucleic acids in which the nucleic acid sequences coding for the polypeptide of the invention and possibly the signal peptide, are recombined with control elements which are heterologous with respect to the ones to which they are normally associated within the bacteria gene and, more particularly, the regulation elements adapted to control their expression in the cellular host which has been chosen for their production.

The invention also relates to recombinant vectors, particularly for cloning and/or expression, comprising a vector sequence, notably of the type plasmid, cosmid or phage, and a recombinant nucleic acid of the invention, in one of the non essential sites for its replication.

According to an advantageous embodiment of the invention, the recombinant vector contains, in one of its non essential sites for its replication, necessary elements to promote the expression of polypeptides according to the invention in a cellular host and possibly a promoter recognized by the polymerase of the cellular host, particularly an inducible promoter and possibly a signal sequence and/or an anchor sequence.

According to another additional embodiment of the invention, the recombinant vector contains the elements enabling the expression by E. coli of a nucleic acid according to the invention inserted in the vector, and particularly the elements enabling the expression of the gene or part thereof of β-galactosidase.

The invention also relates to a cellular host which is transformed by a recombinant vector according to the invention, and comprising the regulation elements enabling the expression of the nucleotide sequence coding for the polypeptide according to the invention in this host.

The invention also relates to a cellular host chosen from among bacteria such as E. coli, transformed by a vector as above defined, or chosen from among eukaryotic organism, such as CHO cells, insect cells, Sf9 cells [Spodoptera frugiperda] infected by the virus Ac NPV (Autographa californica nuclear polyhydrosis virus) containing suitable vectors such as pAc 373 pYM1 or pVC3, BmN [Bombyx mori] infected by the virus BmNPV containing suitable vectors such as pBE520 or p89B310.

The invention relates to an expression product of a nucleic acid expressed by a transformed cellular host according to the invention.

The invention also relates to the use of any secreted polypeptide of the invention as a carrier antigen for foreign epitopes (epitopes of a polypeptide sequence heterologous with respect to the polypeptides of the invention) in the Mycobacterium bovis BCG vaccine strain.

The Mycobacterium bovis BCG vaccine strain used can be available from Institut Pasteur (Paris), under $1173P_2$.

The recombinant DNA comprising the nucleic acid coding for anyone of the polypeptides of the invention and the nucleic acid coding for any foreign epitopes as above defined, can contain the promoter sequence of said polypeptide of the invention, the signal sequence of said polypeptide, possibly the coding part of said polypeptide and the coding nucleic acid of the foreign epitope, said nucleic acid of the foreign epitope being for instance

- either directly located after the signal sequence, and if the coding part of the the polypeptide of the invention is present, upstream the coding part of the polypeptide of the invention,
- or located downstream the coding part of the polypeptide of the invention,
- or located within the coding part of the polypeptide of the invention.

The recombinant DNA as above defined can be transformed into the vaccine strain BCG where it leads to the expression and secretion of a recombinant protein antigen.

From the nucleic acids of the invention, probes (i.e. cloned or synthetic oligonucleotides) can be inferred.

These probes can be from 15 to the maximum number of nucleotides of the selected nucleic acids. The

oligonucleotides can also be used either as amplification primers in the PCR technique (PCR Mullis and Faloona methods in Enzymology, vol. 155, p. 335, 1987) to generate specific enzymatically amplified fragments and/or as probes to detect fragments amplified between bracketing oligonucleotide primers.

The specificity of a PCR assisted hybridization assay can be controlled at different levels.

The amplification process or the detection process or both can be specific. The latter case giving the higher specificity is preferred.

The invention also relates to a process for preparing a polypeptide according to the invention comprising the following steps:

- the culture in an appropriate medium of a cellular host which has previously been transformed by an appropriate vector containing a nucleic acid according to the invention,
- the recovery of the polypeptide produced by the above said transformed cellular host from the above said culture medium, and
- the purification of the polypeptide produced, eventually by means of immobilized metal ion affinity chromatography (IMAC).

The polypeptides of the invention can be prepared according to the classical techniques in the field of peptide synthesis.

The synthesis can be carried out in homogeneous solution or in solid phase.

For instance, the synthesis technique in homogeneous solution which can be used is the one described by Houbenweyl in the book titled "Methode der organischen chemie" (Method of organic chemistry) edited by E. Wunsh, vol. 15-I et II. THIEME, Stuttgart 1974.

The polypeptides of the invention can also be prepared according to the method described by R.D. Merrifield in the article titled "Solid phase peptide synthesis" (J.P. Ham.Socks. , 45, 2149-2154).

The invention also relates to a process for preparing the nucleic acids according to the invention. A suitable method for chemically preparing the single-stranded nucleic acids (containing at most 100 nucleotides of the invention) comprises the following steps :

- DNA synthesis using the automatic β-cyanoethyl phosphoramidite method described in Bioorganic Chemistry 4; 274-325, 1986.

In the case of single-stranded DNA, the material which is obtained at the end of the DNA synthesis can be used as such.

A suitable method for chemically preparing the double-stranded nucleic acids (containing at most 100 bp of the invention) comprises the following steps:

- DNA synthesis of one sense oligonucleotide using the automatic β-cyanoethyl phosphoramidite method described in Bioorganic Chemistry 4; 274-325, 1986, and DNA synthesis of one anti-sense oligonucleotide using said above-mentioned automatic β-cyanoethyl phosphoramidite method,
- combining the sense and anti-sense oligonucleotides by hybridization in order to form a DNA duplex,
- cloning the DNA duplex obtained into a suitable plasmid vector and recovery of the DNA according to classical methods, such as restriction enzyme digestion and agarose gel electrophoresis.

A method for the chemical preparation of nucleic acids of length greater than 100 nucleotides - or bp, in the case of double-stranded nucleic acids - comprises the following steps :

- assembling of chemically synthesized oligonucleotides, provided at their ends with different restriction sites, the sequences of which are compatible with the succession of amino acids in the natural peptide, according to the principle described in Proc. Nat. Acad. Sci. USA 80; 7461-7465, 1983,
- cloning the DNA thereby obtained into a suitable plasmid vector and recovery of the desired nucleic acid according to classical methods, such as restriction enzyme digestion and agarose gel electrophoresis.

The invention also relates to antibodies themselves formed against the polypeptides according to the invention.

It goes without saying that this production is not limited to polyclonal antibodies.

It also relates to any monoclonal antibody produced by any hybridoma liable to be formed according to classical methods from splenic cells of an animal, particularly of a mouse or rat, immunized against the purified polypeptide of the invention on the one hand, and of cells of a myeloma cell line on the other hand, and to be selected by its ability to produce the monoclonal antibodies recognizing the polypeptide which has been initially used for the immunization of the animals.

The invention also relates to any antibody of the invention labeled by an appropriate label of the enzymatic, fluorescent or radioactive type.

The peptides which are advantageously used to produce antibodies, particularly monoclonal antibodies, are the following ones gathered in Table 1 (referring to Figure 1) :

Table 1

38	H_2N-DGLRAQDDYNGWDINTPAFE-COOH	57
78	H_2N-TDWYQPSQSNGQNYTYKWET-COOH	97
174	H_2N-ANSMWGPSSDPAWKRNDPMV-COOH	193
204	H_2N-RIWVYCGNGTPSDLGGDNIP-COOH	223
235	H_2N-NQTFRDTYAADGGRNGVFNF-COOH	254
250	H_2N-GVFNFPPNGTHSWPYWNEQL-COOH	269
275	H_2N-DIQHVLNGATPPAAPAAPAA-COOH	294

The amino acid sequences are given in the 1-letter code.

Variations of the peptides listed in Table 1 are also possible depending on their intended use. For example, if the peptides are to be used to raise antisera, the peptides may be synthesized with an extra cysteine residue added. This extra cysteine residue is preferably added to the amino terminus and facilitates the coupling of the peptide to a carrier protein which is necessary to render the small peptide immunogenic. If the peptide is to be labeled for use in radioimmune assays, it may be advantageous to synthesize the protein with a tyrosine attached to either the amino or carboxyl terminus to facilitate iodination. These peptides possess therefore the primary sequence of the peptides listed in Table 1 but with additional amino acids which do not appear in the primary sequence of the protein and whose sole function is to confer the desired chemical properties to the peptides.

The invention also relates to a process for detecting in vitro antibodies related to tuberculosis in a human biological sample liable to contain them, this process comprising

- contacting the biological sample with a polypeptide or a peptide according to the invention under conditions enabling an in vitro immunological reaction between said polypeptide and the antibodies which are possibly present in the biological sample and
- the in vitro detection of the antigen/antibody complex which may be formed.

Preferably, the biological medium is constituted by a human serum.

The detection can be carried out according to any classical process.

By way of example a preferred method brings into play an immunoenzymatic process according to ELISA technique or immunofluorescent or radioimmunological (RIA) or the equivalent ones.

Thus the invention also relates to any polypeptide according to the invention labeled by an appropriate label of the enzymatic, fluorescent, radioactive... type.

Such a method for detecting in vitro antibodies related to tuberculosis comprises for instance the following steps:

- deposit of determined amounts of a polypeptidic composition according to the invention in the wells of a titration microplate,
- introduction into said wells of increasing dilutions of the serum to be diagnosed,
- incubation of the microplate,
- repeated rinsing of the microplate,
- introduction into the wells of the microplate of labeled antibodies against the blood immunoglobulins,
- the labeling of these antibodies being carried out by means of an enzyme which is selected from among the ones which are able to hydrolyze a substrate by modifying the absorption of the radiation of this latter at least at a given wave length,
- detection by comparing with a control standard of the amount of hydrolyzed substrate.

The invention also relates to a process for detecting and identifying in vitro antigens of M. tuberculosis in a human biological sample liable to contain them, this process comprising: - contacting the biological sample with an appropriate antibody of the invention under conditions enabling an in vitro immunological reaction between said antibody and the antigens of M. tuberculosis which are possibly present in the biological sample and the in vitro detection of the antigen/antibody complex which may be formed.

Preferably, the biological medium is constituted by sputum, pleural effusion liquid, broncho-alveolar washing liquid, urine, biopsy or autopsy material.

Appropriate antibodies are advantageously monoclonal antibodies directed against the peptides which have been mentioned in Table 1.

The invention also relates to an additional method for the in vitro diagnosis of tuberculosis in a patient liable to be infected by Mycobacterium tuberculosis comprising the following steps:

- the possible previous amplification of the amount of the nucleotide sequences according to the invention, liable to be contained in a biological sample taken from said patient by means of a DNA primer set as above defined,
- contacting the above mentioned biological sample with a nucleotide probe of the invention, under conditions enabling the production of an hybridization complex formed between said probe and said nucleotide sequence,
- detecting the above said hybridization complex which has possibly been formed.

To carry out the in vitro diagnostic method for tuberculosis in a patient liable to be infected by Mycobacterium tuberculosis as above defined, the following necessary or kit can be used, said necessary or kit comprising:

- a determined amount of a nucleotide probe of the invention,
- advantageously the appropriate medium for creating an hybridization reaction between the sequence to be detected and the above mentioned probe,
- advantageously, reagents enabling the detection of the hybridization complex which has been formed between the nucleotide sequence and the probe during the hybridization reaction.

The invention also relates to an additional method for the in vitro diagnosis of tuberculosis in a patient liable to be infected by Mycobacterium tuberculosis comprising :

- contacting a biological sample taken from a patient with a polypeptide or a peptide of the invention, under conditions enabling an in vitro immunological reaction between said polypeptide or peptide and the antibodies which are possibly present in the biological sample and
- the in vitro detection of the antigen/antibody complex which has possibly been formed.

To carry out the in vitro diagnostic method for tuberculosis in a patient liable to be infected by Mycobacterium tuberculosis, the following necessary or kit can be used, said necessary or kit comprising:

- a polypeptide or a peptide according to the invention,
- reagents for making a medium appropriate for the immunological reaction to occur,
- reagents enabling to detect the antigen/antibody complex which has been produced by the immunological reaction, said reagents possibly having a label, or being liable to be recognized by a labeled reagent, more particularly in the case where the above mentioned polypeptide or peptide is not labeled.

The invention also relates to an additional method for the in vitro diagnosis of tuberculosis in a patient liable to be infected by M. tuberculosis, comprising the following steps:

- contacting the biological sample with an appropriate antibody of the invention under conditions enabling an in vitro immunological reaction between said antibody and the antigens of M. tuberculosis which are possibly present in the biological sample and - the in vitro detection of the antigen/antibody complex which may be formed.

Appropriate antibodies are advantageously monoclonal antibodies directed against the peptides which have been mentioned in Table 1.

To carry out the in vitro diagnostic method for tuberculosis in a patient liable to be infected by Mycobacterium tuberculosis, the following necessary or kit can be used, said necessary or kit comprising:

- an antibody of the invention,
- reagents for making a medium appropriate for the immunological reaction to occur,
- reagents enabling to detect the antigen/antibody complexes which have been produced by the immunological reaction, said reagent possibly having a label or being liable to be recognized by a label reagent, more particularly in the case where the above mentioned antibody is not labeled.

An advantageous kit for the in vitro diagnosis of tuberculosis comprises:

- at least a suitable solid phase system, e.g. a microtiter-plate for deposition thereon of the biological sample to be diagnosed in vitro,
- a preparation containing one of the monoclonal antibodies of the invention,
- a specific detection system for said monoclonal antibody,
- appropriate buffer solutions for carrying out the immunological reaction between a test sample and said monoclonal antibody on the one hand, and the bonded monoclonal antibodies and the detection system on the other hand.

The invention also relates to a kit, as described above, also containing a preparation of one of the polypeptides or peptides of the invention, said antigen of the invention being either a standard (for quantitative determination of the antigen of M. tuberculosis which is sought) or a competitor, with respect to the antigen which is sought, for the kit to be used in a competition dosage process.

The invention also relates to an immunogenic composition comprising a polypeptide or a peptide according to the invention, in association with a pharmaceutically acceptable vehicle.

The invention also relates to a vaccine composition comprising among other immunogenic principles anyone of the polypeptides or peptides of the invention or the expression product of the invention, possibly coupled to a natural protein or to a synthetic polypeptide having a sufficient molecular weight so that the conjugate is able to induce in vivo the production of antibodies neutralizing Mycobacterium tuberculosis, or induce in vivo a cellular immune response by activating M. tuberculosis antigen-responsive T cells.

The peptides of the invention which are advantageously used as immunogenic principle are the ones mentioned in Table 1.

Other characteristics and advantages of the invention will appear in the following examples and the figures illustrating the invention.

FIGURE LEGENDS

Figure 1 :

Figure 1 represents the nucleotide and amino acid sequence of the 85-C antigen containing region of M. tuberculosis :

The previously identified 28 residue NH_2-terminal amino acid sequence of the mature protein is underlined with a double line. One additional ATG codon, downstream of the ATG at position 150 is underlined. Since the precise length of the signal sequence could not be determined, the option taken here represents the 46 amino acid signal peptide corresponding to ATG_{150}. The putative signal peptide sequence is represented in italic capitals. The top drawing represents the sequencing strategy. Arrows indicate the direction of dideoxy-sequencing either in DNA subcloned as double stranded DNA in Blue Scribe M13+ or as single stranded DNA in the mp18 M13 vector. The entire sequence was determined using synthetic oligonucleotides represented as gray boxes on the figure.

Figure 2 :

Figure 2 represents the homology between known nucleotide and amino acid sequence of the antigen 85 and the 85-C antigen of M. tuberculosis :

A- Comparison of the DNA sequences of antigen 85-A, B and C :

DNA sequences have been aligned with the "Align" program which visualizes multiple alignments. In this presentation, sequence differences are outlined :

(•) indicate identical residues ; (-) indicates a gap ; (any letter) indicates a substitution.

All the sequences are compared and aligned to that of the first line (gene 85-A).

85-A : DNA sequence from M. tuberculosis (Borremans L. et al., 1989, "Cloning, sequence determination and expression of a 32-Kilodalton protein gene of Mycobacterium tuberculosis" Infect. Immun. 57:3123).

85-B : DNA sequence from α-antigen of Mycobacterium bovis (strain Tokyo)(Matsuo K. et al., 1988, "Cloning and expression of the Mycobacterium bovis BCG gene for extracellular α-antigen" J. Bacteriol. 170:3847).

85-C : DNA sequence from antigen 85-C (the present invention).

85-B-Kans : DNA sequence from antigen 85-B from M. kansasii (Matsuo K. et al., 1990, "Cloning and expression of the gene for cross-reactive α antigen of Mycobacterium kansasii" Infect. Immun. 58:550).

85-C-BCG : DNA sequence from Mycobacterium bovis BCG strain 1173P2 (the present invention).

([>)

indicates the presumed initiation codon for each gene.

(↓) indicates the first phenylalanine residue of the mature protein.

(STOP)

indicates the termination codon of each gene.

P78 and P79 are sense and antisense primers used for PCR amplification

85-A, -B, -C sequences used for the synthesis of specific synthetic oligonucleotides probes are framed.

The indicated restriction sites have been used to prepare the three type specific probes.

B- Comparison of the Pre-protein sequences of antigen 85-A, B and C :

DNA sequences have been aligned with the "Align" program which permits multiple alignments. In this presentation, sequence differences are outlined :

(•) indicate identical residues ; (-) indicates a gap ; (any letter) indicates a substitution.

All the sequences are compared and aligned to that of the first line (gene 85-A).

85-A : Protein sequence from M. tuberculosis (Borremans L. et al., 1989, "Cloning, sequence determination and expression of a 32-Kilodalton protein gene of Mycobacterium tuberculosis" Infect. Immun. 57:3123).

85-B : Protein sequence from α-antigen of Mycobacterium bovis (strain Tokyo)(Matsuo K. et al., 1988, "Cloning and expression of the Mycobacterium bovis BCG gene for extracellular α-antigen" J. Bacteriol. 170:3847).

85-C : Protein sequence from antigen 85-C (the present invention).

85-B-Kans : Protein sequence from antigen 85-B from M. kansasii (Matsuo K. et al., 1990, "Cloning and expression of the gene for cross-reactive α antigen of Mycobacterium kansasii" Infect. Immun. 58:550).

85-C-BCG : Protein sequence from Mycobacterium bovis BCG strain 1173P2 (the present invention).

The "C" characteristic motif is framed.

Figure 3 :

Figure 3 represents the hydropathy pattern of the M. tuberculosis 32-kDa (antigen 85-A), the α-antigen of BCG (antigen 85-B) and antigen 85-C from M. tuberculosis, amino acid sequences :

The sequence of the three pre-proteins (including the presumed signal peptide signals) have been analyzed using the Kyte and Doolittle method (Borremans L. et al., 1989, "Cloning, sequence determination and expression of a 32-Kilodalton protein gene of Mycobacterium tuberculosis" Infect. Immun. 57:3123) with a window of eight amino acids. Each bar on the axes represents 50 amino acids. Since the length of signal sequences are slightly different (43, 40 and 46 residues for the three proteins 85-A, 85-B, 85-C) the patterns are aligned to the first residue of the three mature proteins. Plain lines are used to align hydrophobic peaks and a dashed line to align hydrophilic peaks.

Figures 4A and 4B :

Figure 4A represents the restriction endonuclease maps of the three genes 85-A, 85-B and 85-C :

The map of gene 85-A is derived from Borr et al. (Borremans L. et al., 1989, "Cloning, sequence determination and expression of a 32-Kilodalton protein gene of Mycobacterium tuberculosis" Infect. Immun. 57:3123). The map of 85-B was obtained from clone 5.1 derived from our Mycobacterium bovis BCG 1173P2 λgt11 recombinant library (De Wit L. et al., 1990, "Nucleotide sequence of the 32 kDa-protein gene (antigen 85A) of Mycobacterium bovis BCG" Nucl. Ac. Res. 18:3995). For the restriction enzymes used, this map is identical to that published for M. bovis BCG (strain Tokyo) (Matsuo K. et al., 1988, "Cloning and expression of the Mycobacterium bovis BCG gene for extracellular α-antigen" J. Bacteriol. 170:3847). The coding region of the 85-B antigen is positioned according to Matsuo et al. (Matsuo K. et al., 1988, "Cloning and expression of the Mycobacterium bovis BCG gene for extracellular α-antigen" J. Bacteriol. 170:3847).

The map of 85-C corresponds to the restriction map of clone 11.2 that was obtained from the M. tuberculosis λgt11 library from R. Young (Young R.A. et al., 1985, "Dissection of Mycobacterium tuberculosis antigens using recombinant DNA" Proc. Natl. Acad. Sci. USA 82:2583) (Material and Methods). The position of the specific 5' DNA restriction fragment used for Southern analysis is indicated on each map, by a double arrow.

Figure 4B represents the Southern analysis of the total genomic DNA from Mycobacterium bovis BCG - (strain 1173P2) :

15 μg DNA of digested DNA was applied per lane. Hybridization was either with oligonucleotide probes A, B, C (as described in Fig. 2A) or, after deshybridization, with a larger DNA fragment in conditions described in Material and methods. Part 85-C was obtained on a separate gel. Molecular weights of the hybridizing bands were calculated by comparison with standards.

Figure 5 :

Figure 5 represents the pulse field electrophoresis of Mycobacterium tuberculosis DNA :

DNA from three strains of Mycobacterium tuberculosis was digested with DraI and separated by Pulse field electrophoresis on an agarose gel together with a bacteriophage λ DNA 'ladder' as described in Material and methods. After transfer to Nylon filters, hybridization with the three probes 85-A, 85-B, 85-C was as described under Fig. 4. Molecular weights of the hybridizing bands were calculated by comparison with those of the λ DNA 'ladder'.

MATERIAL AND METHODS

1. Preparation of genomic DNA (Thole J. et al., 1985, "Cloning of Mycobacterium bovis BCG DNA and expression of antigens in Escherichia coli" Infect. Immun. 50:3800) :

M. bovis BCG was cultivated at 37 ° C in Sauton medium and harvested after an additional incubation of 18 hrs in the presence of 1% glycine added at the end of the late exponential growth phase. The bacteria were treated with lysozyme and proteinase K, lysed with sodium dodecyl sulfate, phenol extracted and ethanol precipitated.

2. Genomic libraries :

A λgt11 recombinant library constructed from genomic DNA of M. tuberculosis (Erdman strain), was obtained from Young R.A. et al., 1985, "Dissection of Mycobacterium tuberculosis antigens using recombinant DNA" Proc. Natl. Acad. Sci. USA 82:2583.

A second λgt11 recombinant library was prepared with genomic DNA from M. bovis BCG (De Wit L. et al., 1990, "Nucleotide sequence of the 32 kDa-protein gene (antigen 85A) of Mycobacterium bovis BCG" Nucl. Ac. Res. 18:3995).

3. Oligonucleotides :

Oligonucleotides were synthesized on an Applied Biosystems DNA synthesizer model 381A, purified on OPC-cartridges (Applied Biosystems), lyophilized and dissolved in TE buffer (10 mM Tris-HCl, pH 7,4).

^{32}P labeling of the oligonucleotides was as described in Sambrook J. et al., 1989, "Molecular cloning:a laboratory manual" Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

4. PCR :

4.1. Amplification and cloning of genomic DNA:

50 ng of Mycobacterium bovis BCG DNA was amplified in a 50 µl reaction containing 1xPCR-buffer (Amersham), 200 µM dNTP, 1 µM each of sense P78 (5'-CCGGAATTCATGGGCCGTGACATCAAG) and antisense P79 (5'-CCGGAATTCGGTCTCCCACTTGTAAGT) oligonucleotide primers (the location of these two primers is indicated in Figure 2A. Both oligonucleotides were added an EcoRI sequence preceded by 3 additional nucleotides), and 2 units of Tag DNA polymerase. After denaturation for 90 seconds at 94 ° C the reaction was submitted to 40 cycles consisting of 1 minute at 93 ° C (denaturation), 90 seconds at 55 ° C (annealing), 2 minutes at 72 ° C (extension), followed by a 5 minute final extension at 72 ° C. After extraction with 150 µl chloroform, the amplified DNA, was washed three times with 0.75 ml H_2O in a Centricon-30 for 6 minutes at 6500 rpm in the Sorvall SS 34 rotor. After digestion with EcoRI the DNA was ligated into EcoRI-digested, phosphatase-treated Bluescribe-M13+ vector. DH5α E. coli (Gibco-BRL) were transformed and plated on Hybond-N filters. Colonies were selected by hybridization with ^{32}P-oligonucleotide probe-A (5'-TCGCCCGCCCTGTACCTG) and probe-B (5'-TCACCTGCGGTTTATCTG). Hybridization and washing conditions for the oligonucleotides were as described by Jacobs et al. (Jacobs et al., 1988, "The thermal stability of oligonucleotide duplexes is sequence independent in tetraalkylammonium salt solutions: application to identifying recombinant DNA clones" Nucl. Ac. Res. 16:4637).

4.2. Amplification of λgt11 plaques DNA :

10 µl of each λgt11 plaque (1 plaque was resuspended in 1 ml SM medium containing 5% chloroform) were amplified in a 100 µl reaction containing 200 µM of each dNTP, 1 µM each of sense oligonucleotide B and antisense oligonucleotide P79 primers, and 1x PCR buffer (Amersham). After an initial denaturation for

90 seconds at 94° C, 2 units of Taq DNA polymerase were added. The reaction was submitted to 40 cycles consisting of 1 minute at 94° C (denaturation), 90 seconds at 60° C (annealing), 2 minutes at 72° C (extension) followed by a 5 minutes final extension at 72° C. 10 µl were analyzed on a 2% agarose gel stained with ethidium bromide.

5. Screening of the λgt11 M. tuberculosis and Mycobacterium bovis BCG recombinant DNA libraries:

The two λgt11 recombinant libraries were screened by colony hybridization (Sambrook J. et al., 1989, "Molecular cloning:a laboratory manual" Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.) with a 800 bp Hind III fragment which does not discriminate gene 85-A from 85-B (see Fig. 2A). 12 positive M. tuberculosis and 12 Mycobacterium bovis BCG plaques were retained and screened by PCR-amplification with primers B (sense) and P79 (antisense). Non-A recombinant λgt11 recombinants were further screened by hybridization with ^{32}P-labeled-C oligonucleotide-probe (5'-TCGCAGAGCAACGGGCCAGAACTAC) as described above.

From the M. tuberculosis λgt11 library, one selected bacteriophage #11 was partially digested with EcoRI and its 5 kbp insert was subcloned in Bluescribe-M13+. From this recombinant plasmid named 11-2, a 3,500 bp BamHI-EcoRI fragment was subcloned in M13-mp18 and M13-mp19 (Sambrook J. et al., 1989, "Molecular cloning:a laboratory manual" Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.).

6. Recombinant DNA analysis :

It was as described in Borremans L. et al., 1989, "Cloning, sequence determination and expression of a 32-Kilodalton protein gene of Mycobacterium tuberculosis" Infect. Immun. 57:3123.

7. Sequencing :

Sequence analysis was done by the primer extension dideoxy termination method of Sanger et al. (Sanger F. et al., 1977, "DNA sequencing with chain termination inhibitors" Proc. Natl. Acad. Sci. USA 74:5463) after subcloning of specific fragments in Bluescribe-M13+ (Chen E.J. et al., 1985, "Supercoil sequencing: a fast simple method for sequencing plasmid DNA" DNA 4:165) or in mp18 and mp19 M13 vectors. Sequence analysis was greatly hampered by the high GC content of the M. tuberculosis DNA (65%). Sequencing reactions were therefore performed with several DNA polymerases according to manufacturers protocols : T7 DNA polymerase ("Sequenase" USB), T7 DNA polymerase (Pharmacia), and Taq DNA polymerase (Promega) using 7-deaza-dGTP instead of dGTP. Several oligodeoxynucleotides were synthesized and used to focus on ambiguous regions of the sequence. The sequencing strategy is summarized in Fig. 1.

8. Sequence comparison and analysis :

Routine computer aided analysis of the nucleic acid and deduced amino acid sequences were performed with the LGBC program from Bellon B., 1988, "Apple Macintosh programs for nucleic and protein sequence analysis" Nucleic Acid Res. 16:1837. Homology searches used the FASTA programs from Pearson W.R. et al., 1988, "Improved tools for biological sequence comparison" Proc. Natl. Acad. Sci. USA 85:2444, and the various DNA and protein data bank from the EMBL-server facilities. Multiple alignments were obtained with 'Align 1.01' (Scientific and Educational Software).

9. Southern blot analysis :

Genomic DNA from Mycobacterium bovis BCG was completely digested with SphI, EcoRI or KpnI, electrophoresed on a 1% agarose gel, transferred to Hybond-N filter (Amersham) after denaturation and neutralization and either hybridized with ^{32}P-labeled-oligonucleotide probes (A, B, C) in the conditions described in Jacobs et al., 1988, "The thermal stability of oligonucleotide duplexes is sequence independent in tetraalkylammonium salt solutions: application to identifying recombinant DNA clones" Nucl. Ac. Res. 16:4637, or random-primed ^{32}P-labeled DNA restriction fragments that were found to discriminate the 3 genes 85-A, 85-B, and 85-C.

Probes 85-A was a 230 bp PstI fragment from plasmid BY-5 (Borremans L. et al., 1989, "Cloning, sequence determination and expression of a 32-Kilodalton protein gene of Mycobacterium tuberculosis" Infect. Immun. 57:3123 and Fig. 2A). Probes 85-B was a 400 bp SmaI-EcoRV fragment from a 85-B

recombinant plasmid named 5.1, derived from our Mycobacterium bovis BCG λgt11 library, whose map is presented in Fig. 4A (see also Fig. 2A). Probe 85-C was a 280 bp SmaI-KpnI fragment from plasmid 11.2 (see also Fig. 4A and 2A).

These DNA fragments were prepared by gel electrophoresis on low melting point agarose followed by a rapid purification on Qiagen (marketed by : Westburg, Netherlands) (tip 5) according to manufacturers protocol and labeled in the presence of α-^{32}P-dCTP (Feinberg A.P. et al., 1983, "A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity" Anal. Biochem. 132:6).

10. Pulse Field electrophoresis DNA separation :

DNA preparation, restriction enzyme digestion and pulse-field gel electrophoresis were performed as previously described (Vincent Levy-Frebault V. et al., 1990, "DNA polymorphism in Mycobacterium paratuberculosis, "wood pigeon mycobacteria" and related mycobacteria analyzed by field inversion gel electrophoresis", J. Clin. Microbiol. 27:2723). Briefly cells from fresh cultures were mixed with 1% low-melting-point agarose (v/v) and submitted to successive treatments with zymolase (Seikagaki Kogyo, Tokyo, Japan), lysozyme, and sodium dodecyl sulfate in the presence of proteinase K (Boehringer GmbH, Mannheim, Germany). After inactivation of proteinase K with phenylmethylsulfonyl fluoride (Bio-Rad Laboratories), agarose blocks were digested overnight with 50 U of DraI (Bio-Rad Laboratories). Then blocks were loaded into a 1% agarose gel prepared and electrophoresed in 0.66 TBE (Tris-boric acid - EDTA) (Vincent Levy-Frebault V. et al., 1990, "DNA polymorphism in Mycobacterium paratuberculosis, "wood pigeon mycobacteria" and related mycobacteria analyzed by field inversion gel electrophoresis", J. Clin. Microbiol. 27:2723). Field inversion gel electrophoresis was carried out using a Dnastar Pulse (Dnastar, USA) apparatus. Forward and reverses pulses were set at 0.33 s and 0.11 s at the beginning of the run and 60 s and 20 s (or 30 s and 10 s) at the end of the run depending on the molecular weight zone to be expanded. The run time was set at 36 h, the voltage used was 100 V and producing about 325 mA and temperature was maintained at 18 °C. Lambda concatemers were used as molecular weight markers. At the end of the run, the gels were stained with ethidium bromide, photographed under UV light and transferred onto Nylon membranes according to Maniatis T. et al., 1982, "Molecular cloning: a laboratory manual" Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. 545 pp.

RESULTS

1. Cloning of the 85-C gene of M. tuberculosis :

Since no specific probe or monoclonal antibody was available to detect specifically an 85-C or related antigen which was expected to bear extensive homology to gene 85-A and gene 85-B this screening required the development of a new procedure. The strategy used was based on the PCR amplification of a 245 bp DNA fragment coding for amino acids 18-98 of the mature antigen 85-A chosen because it is surrounded at both ends by highly conserved DNA sequences when the sequences of antigen A and B are aligned (see primers P78 and P79 in Fig. 2A). It was thus speculated that an equivalent homology might exist with the sequence of antigen 85-C in the same region.

From Mycobacterium bovis BCG genomic DNA a 245 bp DNA fragment was readily obtained. The latter was purified and subcloned in a Bluescribe M13+ vector after digestion with EcoRI. About 80 recombinant plasmid containing colonies were tested by plating on nylon filters and hybridized in stringent conditions with a labeled synthetic oligonucleotide recognizing either sequence 85-A (5'-TCGCCCGCCCTGTACCTG) or sequence B (5'-TCACCTGCGGTTTATCTG) within the PCR amplified fragment (see Fig. 2A). Several clones that hybridized with each probe were sequences and the sequences were all identical to sequence 85-A in the clones hybridizing with probe A. In those hybridizing with probes 85-B, two kinds of sequences were found : either the 85-B sequence (Matsuo K. et al., 1988, "Cloning and expression of the Mycobacterium bovis BCG gene for extracellular α-antigen" J. Bacteriol. 170:3847) or a new 85-B-related sequence described in Fig. 2A. The latter presents a marked sequence divergence covering 24 nucleotides which is totally distinct from sequence A and B (Fig. 2) (The homology to sequence B is only 40% in this region). Assuming these inserts might represent an amplified fragment of the 85-C gene and that this 24 nucleotide sequence is characteristic of the putative 85-C gene an oligonucleotide probe (probe 85-C) based on this sequence was synthesized.

The latter probe was labeled with ^{32}P and used to screen a collection of 24 λgt11 recombinant phages that were selected in our M. tuberculosis and Mycobacterium bovis BCG λgt11 libraries by hybridization with a 800 bp HindIII DNA fragment of the previously cloned gene 85-A (Borremans L. et al., 1989,

"Cloning, sequence determination and expression of a 32-Kilodalton protein gene of Mycobacterium tuberculosis" Infect. Immun. 57:3123).

Among those phages, those containing the "B" oligonucleotide sequence were further selected by analytical PCR assay using probe B (sense) and oligonucleotide P79 as antisense. This small collection of "85-B-related" λgt11 recombinant was then hybridized with the described oligonucleotide 85-C and one hybridizing λgt11-M. tuberculosis recombinant was retained, characterized by restriction mapping and sequenced.

2. Sequence of the 85-C gene :

The 1211 nucleotide sequence derived from various sequenced fragments is represented in Fig. 1. The DNA sequence contains a 1,020 bp long open reading frame, starting at position 150 and ending with a TGA codon at position 1170. The common NH2 terminal amino acid sequence of the antigen 85 proteins, Phe-Ser, Arg-Pro-Gly-Leu (De Bruyn J. et al., 1987, "Purification, partial characterization and identification of a 32 kDa protein antigen of Mycobacterium bovis BCG" Microb. Pathogen. 2:351) could be located within this open reading frame from the nucleotide sequence beginning with a TTC codon at position 288 (Fig. 1). Therefore the DNA region upstream of this sequence is expected to code for a signal peptide required for the secretion of this antigen. The mature protein consists of 294 amino acid residues corresponding to a calculated molecular weight of 32,021.

Interestingly, the N-terminal sequence of the mature protein contains the entire 26 amino acid sequence (phe-ser-arg-pro-gly-leu-pro-val-glu-tyr-leu-gln-val-pro-ser-ala-ser-met-gly-arg-asp-ile-lys-val-gln-phe) described by Wiker H.G. et al., 1990, "Evidence for three separate genes encoding the proteins of the mycobacterial antigen 85 complex" Infect. Immun. 58:272, and which differs only from the common 85-B and 85-A sequence by an alanine instead of a proline in position 16 of the mature protein. Two ATG codons were found to precede the TTC phenylalanine codon at nucleotide position 288 (Fig. 1) in the same reading frame. Use of these two ATG would lead to the synthesis of signal peptides of either 21 or 46 amino acid residues (the latter situation has been represented in Fig. 1 for reasons indicated below).

The base composition of antigen 85-C gene was identical to that of the 85-A gene with an overall G-C composition of 64,57% and a strong preference for G or C in codon position 3 (average 85%). In contrast to antigen 85-A and 85-B that contain 3 cysteins, the sequence of antigen 85-C shows a single cystein residue at position 254. In fact the two substituted cysteins are located in the region of the mature 85-C protein which contains the largest divergent sequence bloc (Fig. 2B) (SQSNGQNY) (The corresponding DNA sequence was used to synthesize the oligonucleotide probe "C" (see above)). Not surprisingly, this hydrophilic region is also the most divergent when the hydropathy plots of the 3 antigens are compared and thus could be either a variable "epitope" of all 85-antigens and/or a characteristic epitope of antigen 85-C since it was also found in antigen 85-C from M. bovis BCG.

Another characteristic feature of antigen 85-C is the presence of the unusual hydrophobic repetitive proline alanine motive PPAAPAAPAA at the carboxy-terminal of the molecule.

3. Hydropathy pattern :

The hydropathy pattern of M. tuberculosis 85-C antigen was determined by the method of Kyte and Doolittle (Kyte J. et al., 1982, "Simple method for displaying the hydropathy character of a protein" J. Mol. Biol. 157:105). The octapeptide profiles were compared to antigen 85-A and 85-B (Fig. 3). As anticipated from the amino acid sequences, the patterns are roughly similar for the three antigens except for some major differences at region 84-92 and in the carboxy-terminal part of the three proteins.

4. Sequence homologies :

DNA sequences from antigen 85-A (Borremans L. et al., 1989, "Cloning, sequence determination and expression of a 32-Kilodalton protein gene of Mycobacterium tuberculosis" Infect. Immun. 57:3123 ; De Wit L. et al., 1990, "Nucleotide sequence of the 32 kDa-protein gene (antigen 85A) of Mycobacterium bovis BCG" Nucl. Ac. Res. 18:3995), 85-B (Matsuo K. et al., 1988, "Cloning and expression of the Mycobacterium bovis BCG gene for extracellular α-antigen" J. Bacteriol. 170:3847 ; Maniatis T. et al., 1982, "Molecular cloning: a laboratory manual" Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. 545 pp) and 85-C were aligned. An alignment of the three DNA sequences is shown in Fig. 2A. At the DNA level, the homology is maximal between the regions coding for the 3 mature proteins. In this region, the homology between A and B is 77.5% whereas it reaches only 70.8% between the coding regions of genes A and C

and 71.9% between B and C, respectively. Beyond nucleotide 1369 of sequence 85-A and upstream nucleotide position 475 (i.e. within the signal sequence and promoter region) there is practically no homology between the 3 sequences. No significant homology was detected to other DNA sequences present in latest release of GenBank-EMBL.

Homologies at the amino acid level, are presented in the alignment (Fig. 2B) and summarized in Table 1, indicating again a higher homology between sequences A and B (80.4%) than between B/C or A/C.

Other comparisons between the 85-C antigen and the entire SwissProt-NBRF data bank failed to detect any significant homologies to the 85-C antigen amino acid sequence. As for the 85-A antigen, the 85-C sequence does not contain the RGD motif of fibronectin binding proteins nor does it share any homology to the known fibronectin receptors or to the fibronectin binding protein from Staphylococcus aureus.

Comparison of the partial DNA sequences of the 85-C gene of the M. bovis BCG strain 1173P2 with that of the M. tuberculosis strain shows 93.9% homology at the DNA level (Fig. 2A). The region corresponding to the oligo C is entirely conserved. Major differences in the region of the 85-C gene DNA sequence corresponding to nucleotides 615 to 629 of antigen 85-A result in 5 amino acid changes (Fig. 28). Interestingly the DNA sequence of BCG 85-C within this divergent region is 100% homologous to the sequence 85-8 (Fig. 2A).

5. Genome characterization :

In order to confirm the existence of different genes coding for the antigen 85 complex M. bovis BCG genomic DNA was digested with SphI, EcoRI and KpnI and the distribution of radioactive signals was examined in Southern blot after hybridization with three specific oligonucleotide (A, B, C) probes (see Material and Methods and Fig. 2A). Three clearly distinct patterns were obtained confirming the specificity of these probes. Similar type specific profiles could be obtained with three random priming labeled DNA restriction fragments (probe 85-A, 230 bp, 85-B, 400 bp, 85-C 280 bp) which were selected within the promoter signal sequence of the three DNAs (Fig. 2A and 4A). With these three DNA restriction fragments, additional weak bands are also observed which clearly correspond to cross hybridization of the probes to the other two genes. With probe 85-C an additional KpnI fragment was observed that does not hybridize to the C-oligonucleotide probe. This probably indicates that the corresponding KpnI site is located upstream of this gene. Furthermore the size of the observed restriction fragments are not always exactly as expected from the restriction maps of the corresponding cloned genes. These discrepancies probably correspond to some minor sequence differences (restriction polymorphism) possibly in non coding DNA regions (outside of the DNA coding for the antigen 85) between strain of M. bovis BCG and the M. bovis BCG (strain Tokyo) and M. tuberculosis respectively.

6. Pulse field analysis of M. tuberculosis genomic DNA :

When the largest available 85-A clone BY-5 was hybridized (Fig. 4A) with oligonucleotide B, no positive signal was detected whereas oligonucleotide A gave a positive hybridization (not shown). This indicates that gene B is not located within 2-2.5 kb of the 5' and 4.0 kb of the 3' border of gene A (Fig. 4A). To confirm and extent this result, pulse-field separated DraI-digested M. tuberculosis genomic DNA was further hybridized with three specific DNA probes 85-A, 85-B and 85-C in stringent conditions.

Eight strains of M. tuberculosis were compared showing six different patterns, three of which are illustrated in Fig. 5. For most strains examined the three probes hybridized to fragments of different sizes. For instance, in M. tuberculosis H37Ra, the size of the DraI fragments hybridizing with probes 85-A, B and C were about 242 kb, 212 kb and 245 kb for strain H37Ra, 403 kb, 212 kb and 104 kb for strain H37Rv and 355 kb, 104 kb and 153 kb for strain "1025". Although various strains show some restriction fragment length polymorphism with restriction endonuclease DraI, the simplest interpretation of these results is that the three antigen 85 genes are distantly located (> 100 kb) within the mycobacterial genome.

Claims

1. Nucleic acid

* containing a nucleotide sequence extending from the extremity constituted by the nucleotide at position (1) to the extremity constituted by the nucleotide at position (149) represented on Figure 1,
* or containing one at least of the nucleotide sequences coding for the following peptides or polypeptides :

- the one extending from the extremity constituted by amino acid at position (-46) to the extremity constituted by amino acid at position (-1) represented on Figure 1, or
- the one extending from the extremity constituted by amino acid at position (-21) to the extremity constituted by amino acid at position (-1) represented on Figure 1, or
- SQSNGQNY, or
- PMVQIPRLVA, or
- GLTLRTNQTPRDTYAADGGRNG, or
- PPAAPAAPAA,

* or containing nucleotidic sequences :
- hybridizing with the above-mentioned nucleotide sequences, or their complements,
- complementary to the above-mentioned nucleotide sequences, or
- which are the above-mentioned nucleotide sequences wherein T can be replaced by U.

2. Nucleic acid according to claim 1, containing the nucleotide sequence coding for the following peptide :
SQSNGQNY
and liable to hybridize with the following nucleotide sequence :
CGGCTGGGAC(or T)ATCAACACCCCGGC
and liable to hybridize neither with
GCCTGCGGCAAGGCCGGTTGCCAG
nor with
GCCTGCGGTAAGGCTGGCTGCCAG
nor with
GCCTGCGGCAAGGCCGGCTGCACG.

3. Nucleic acid according to anyone of claims 1 or 2, containing an open reading frame coding for a polypeptide
- liable to react selectively with human sera from tuberculosis patients and particularly patients developing an evolutive tuberculosis,
- or liable to be recognized by antibodies also recognizing the amino acid sequence extending from the extremity constituted by amino acid at position (1) to the extremity constituted by amino acid at position (294) represented on Figure 1,
- or liable to generate antibodies recognizing the amino acid sequence extending from the extremity constituted by amino acid at position (1) to the extremity constituted by amino acid at position (294) represented on Figure 1.

4. Nucleic acid according to anyone of claims 1 to 3, containing an open reading frame and coding for a mature polypeptide of about 30 to about 35 kD and containing a sequence coding for a signal sequence.

5. Nucleic acid according to claim 1, containing a nucleic acid coding for the following polypeptides :
- the one extending from the extremity constituted by amino acid at position (-46) to the extremity constituted by amino acid at position (-1) represented on Figure 1, or
- the one extending from the extremity constituted by amino acid at position (-21) to the extremity constituted by amino acid at position (-1) represented on Figure 1, or
- the one extending from the extremity constituted by amino acid at position (-46) to the extremity constituted by amino acid at position (294) represented on Figure 1, or
- the one extending from the extremity constituted by amino acid at position (-21) to the extremity constituted by amino acid at position (294) represented on Figure 1, or
- the one extending from the extremity constituted by amino acid at position (1) to the extremity constituted by amino acid at position (294) represented on Figure 1.

6. Nucleic acid sequence according to anyone of claims 1 or 2, containing a nucleic acid coding for a polypeptide sequence extending from the extremity constituted by amino acid at position (18) to the extremity constituted by amino acid at position (99) represented on Figure 2B, on the fifth line.

7. Polypeptide coded by the nucleic acids of anyone of claims 1 to 6.

8. Polypeptide according to claim 7, containing in its polypeptide chain one at least of the following amino

acid sequences :

- the one extending from the extremity constituted by amino acid at position (-46) to the extremity constituted by amino acid at position (-1) represented on Figure 1,
- or the one extending from the extremity constituted by amino acid at position (-21) to the extremity constituted by amino acid at position (-1) represented on Figure 1,
- SQSNGQNY,
- PMVQIPRLVA,
- GLTLRTNQTPRDTYAADGGRNG,
- PPAAPAAPAA.

9. Polypeptide according to claim 8, containing in its polypeptide chain the following amino acid sequence:
SQSNGQNY
and the amino acid sequence
GWDINTPA
and containing not the amino acid sequence
ACGKAGCQ
and not the amino acid sequence
ACGKAGCT

10. Polypeptide according to anyone of claims 8 or 9, liable to react selectively with human sera from tuberculosis patients and particularly patients developing an evolutive tuberculosis,
or liable to be recognized by antibodies also recognizing the polypeptidic sequence extending from the extremity constituted by amino acid at position (1) to the extremity constituted by amino acid at position (294) represented on Figure 1,
or liable to generate antibodies recognizing the polypeptidic sequence extending from the extremity constituted by amino acid at position (1) to the extremity constituted by amino acid at position (294) represented on Figure 1.

11. Polypeptide according to anyone of claims 8 to 10, of about 30 to about 35 kD and preceded by a signal peptide.

12. Polypeptide according to claim 8, containing in its polypeptide chain one at least of the following amino acid sequences :

- the one extending from the extremity constituted by amino acid at position (1) to the extremity constituted by amino acid at position (294) represented on Figure 1, or
- the one extending from the extremity constituted by amino acid at position (-46) to the extremity constituted by amino acid at position (294) represented on Figure 1, or
- the one extending from the extremity constituted by amino acid at position (-21) to the extremity constituted by amino acid at position (294) represented on Figure 1, or
- the one extending from the extremity constituted by amino acid at position (-46) to the extremity constituted by amino acid at position (-1) represented on Figure 1, or
- the one extending from the extremity constituted by amino acid at position (-21) to the extremity constituted by amino acid at position (-1) represented on Figure 1.

13. Polypeptide according to anyone of claims 8 or 9, containing in its amino polypeptide chain the amino acid sequence extending from the extremity constituted by amino acid at position (18) to the extremity constituted by amino acid at position (99) represented on Figure 2B, on the fifth line.

14. Amino acid sequences constituted by a polypeptide according to anyone of claims 7 to 13, and a protein or an heterologous sequence with respect to said polypeptide, said protein or heterologous sequence comprising from about 1 to about 1.000 amino acids, the heterologous protein being advantageously β-galactosidase.

15. Recombinant nucleic acid containing at least one of the nucleotide sequences according to anyone of claims 1 to 6, inserted in a heterologous nucleic acid.

16. Recombinant vector, particularly for cloning and/or expression, comprising a vector sequence, notably

of the type plasmid, cosmid or phage, and a recombinant nucleic acid according to anyone of claims 1 to 6, in one of the non essential sites for its replication.

17. Recombinant vector according to claim 16, containing in one of its non essential sites for its replication necessary elements to promote the expression of polypeptides according to anyone of claims 7 to 13 in a cellular host and possibly a promoter recognized by the polymerase of the cellular host, particularly an inductible promoter and possibly a signal sequence and/or an anchoring sequence.

18. Recombinant vector according to claim 17, containing the elements enabling the expression by E. coli of a nucleic acid according to anyone of claims 1 to 15 inserted in the vector, and particularly the elements enabling the expression of the gene or part thereof of β-galactosidase.

19. Cellular host which is transformed by a recombinant vector according to anyone of claims 16 to 18, and comprising the regulation elements enabling the expression of the nucleotide sequence coding for the polypeptide according to anyone of claims 7 to 13 in this host.

20. Cellular host according to claim 19, chosen from among bacteria such as E. coli, transformed by the vector according to claim 16, or chosen from among eukaryotic organism, transformed by the vector according to claim 16.

21. Expression product of a nucleic acid expressed by a transformed cellular host according to anyone of claims 19 or 20.

22. Antibody characterized by the fact that it is directed against a recombinant polypeptide according to anyone of claims 7 to 13.

23. Nucleotidic probes, hybridizing with anyone of the nucleic acids according to anyone of claims 1 to 6 or with their complementary sequences.

24. Process for preparing a recombinant polypeptide according to anyone of claims 7 to 13 comprising the following steps:

- the culture in an appropriate medium of a cellular host which has previously been transformed by an appropriate vector containing a nucleic acid according to anyone of claims 1 to 6, and
- the recovery of the polypeptide produced by the above said transformed cellular host from the above said culture medium.

25. Method for the in vitro diagnosis of tuberculosis in a patient liable to be infected by Mycobacterium tuberculosis comprising

- contacting a biological sample taken from a patient with a polypeptide according to anyone of claims 7 to 13, under conditions enabling an in vitro immunological reaction between said polypeptide and the antibodies which are possibly present in the biological sample and
- the in vitro detection of the antigen/antibody complex which has been possibly formed.

26. Method for the in vitro diagnosis of tuberculosis in a patient liable to be infected by M. tuberculosis, comprising the following steps :

- contacting the biological sample with an appropriate antibody according to claim 22, under conditions enabling an in vitro immunological reaction between said antibody and the antigens of M. tuberculosis which are possibly present in the biological sample and
- the in vitro detection of the antigen/antibody complex which may be formed.

27. Necessary or kit for an in vitro diagnostic method for tuberculosis in a patient liable to be infected by Mycobacterium tuberculosis according to claim 25, comprising

- a polypeptide according to anyone of claims 7 to 13,
- reagents for making a medium appropriate for the immunological reaction to occur,
- reagents enabling to detect the antigen/antibody complex which has been produced by the immunological reaction, said reagents possibly having a label, or being liable to be recognized by a labeled reagent, more particularly in the case where the above mentioned polypeptide is not labeled.

28. Necessary or kit for an in vitro diagnostic method for tuberculosis in a patient liable to be infected by Mycobacterium tuberculosis according to claim 26, comprising

- an antibody according to claim 22,
- reagents for making a medium appropriate for the immunological reaction to occur,
- reagents enabling to detect the antigen/antibody complexes which have been produced by the immunological reaction, said reagents possibly having a label or being liable to be recognized by a label reagent, more particularly in the case where the above mentioned antibody is not labeled.

29. Immunogenic composition comprising a polypeptide according to anyone of claims 7 to 13, in association with a pharmaceutically acceptable vehicle.

30. Vaccine composition comprising among other immunogenic principles anyone of the polypeptides according to claims 7 to 13 or the expression product of claim 21, possibly coupled to a natural protein or to a synthetic polypeptide having a sufficient molecular weight so that the conjugate is able to induce in vivo the production of antibodies neutralizing Mycobacterium tuberculosis, or induce in vivo a cellular immune response by activating M. tuberculosis antigen-responsive T cells.

31. Peptides of claim 7, advantageously used to produce antibodies, particularly monoclonal antibodies and which have the following amino acid sequences (referring to Figure 1) :

38	H_2N-DGLRAQDDYNGWDINTPAFE-COOH	57
78	H_2N-TDWYQPSQSNGQNYTYKWET-COOH	97
174	H_2N-ANSMWGPSSDPAWKRNDPMV-COOH	193
204	H_2N-RIWVYCGNGTPSDLGGDNIP-COOH	223
235	H_2N-NQTFRDTYAADGGRNGVFNF-COOH	254
250	H_2N-GVFNFPPNGTHSWPYWNEQL-COOH	269
275	H_2N-DIQHVLNGATPPAAPAAPAA-COOH	294

32. Mycobacterium bovis BCG vaccine strain transformed by a recombinant DNA sequence comprising a polypeptide according to anyone of claims 7 to 13, and an epitope of a polypeptide sequence heterologous with respect to said polypeptide.

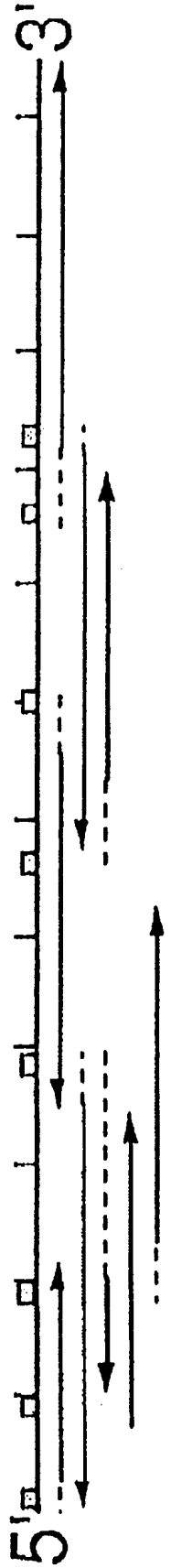


Figure 1

```
                                                                      1
                                                                      AGGTGTCCG
GGCCGACGCTGAATCGTTAGCCAACCGCGATCTCGCGCTGCGGCCACGACATTCGAACTGAGCGTCCTCG
GTGTGTTTCACTCGCCCAGAACAGATTCGACCGCGTCGTGCGCAGATGAGAGTTGGGATTGGTAGTAGCT

-46
MET THR PHE PHE GLU GLN VAL ARG ARG LEU ARG SER ALA ALA THR THR LEU PRO
ATG ACG TTC TTC GAA CAG GTG CGA AGG TTG CGG AGC GCA GCG ACA ACC CTG CCG
        158         167         176         185         194         203
-28
ARG ARG VAL ALA ILE ALA ALA MET GLY ALA VAL LEU VAL TYR GLY LEU VAL GLY
CGC CGC GTG GCT ATC GCG GCT ATG GGG GCT GTC CTG GTT TAC GGT CTG GTC GGT
        212         221         230         239         248         257

-10                                     -1  +1
THR PHE GLY GLY PRO ALA THR ALA GLY ALA Phe Ser Arg Pro Gly Leu Pro Val
ACC TTC GGC GGG CCG GCC ACC GCG GGC GCA TTC TCT AGG CCC GGT CTT CCA GTG
        266         275         284         293         302         311
 9
Glu Tyr Leu Gln Val Pro Ser Ala Ser Met Gly Arg Asp Ile Lys Val Gln Phe
GAA TAT CTG CAG GTG CCA TCC GCG TCG ATG GGC CGC GAC ATC AAG GTC CAG TTC
        320         329         338         347         356         365
 27
Gln Gly Gly Gly Pro His Ala Val Tyr Leu Leu Asp Gly Leu Arg Ala Gln Asp
CAG GGC GGC GGA CCG CAC GCG GTC TAC CTG CTC GAC GGT CTG CGG GCC CAG GAT
        374         383         392         401         410         419
```

Figure 1 (con't 1)

EP 0 499 003 A1

45

Asp	Tyr	Asn	Gly	Trp	Asp	Ile	Asn	Thr	Pro	Ala	Phe	Glu	Glu	Tyr	Tyr	Gln	Ser
GAC	TAC	AAC	GGC	TGG	GAC	ATC	AAC	ACC	CCG	GCC	TTC	GAG	GAG	TAC	TAC	CAG	TCA
		428			437			446			455			464			473

63

Gly	Leu	Ser	Val	Ile	Met	Pro	Val	Gly	Gly	Gln	Ser	Ser	Phe	Tyr	Thr	Asp	Trp
GGG	TTG	TCG	GTG	ATC	ATG	CCC	GTG	GGC	GGC	CAA	TCC	AGT	TTC	TAC	ACC	GAC	TGG
		482			491			500			509			518			527

72

Tyr	Gln	Pro	Ser	Gln	Ser	Asn	Gly	Gln	Asn	Tyr	Thr	Tyr	Lys	Trp	Glu	Thr	Phe
TAT	CAG	CCC	TCG	CAG	AGC	AAC	GGC	CAG	AAC	TAC	ACC	TAC	AAG	TGG	GAG	ACC	TTC
		536			545			554			563			572			581

90

Leu	Thr	Arg	Glu	Met	Pro	Ala	Trp	Leu	Gln	Ala	Asn	Lys	Gly	Val	Ser	Pro	Thr
CTT	ACC	AGA	GAG	ATG	CCC	GCC	TGG	CTA	CAG	GCC	AAC	AAG	GGC	GTG	TCC	CCG	ACA
		590			599			608			617			626			635

108

Gly	Asn	Ala	Ala	Val	Gly	Leu	Ser	Met	Ser	Gly	Gly	Ser	Ala	Leu	Ile	Leu	Ala
GGC	AAC	GCG	GCG	GTG	GGT	CTT	TCG	ATG	TCG	GGC	GGT	TCC	GCG	CTG	ATC	CTG	GCC
		644			653			662			671			680			689

126

Ala	Tyr	Tyr	Pro	Gln	Gln	Phe	Pro	Tyr	Ala	Ala	Ser	Leu	Ser	Gly	Phe	Leu	Asn
GCG	TAC	TAC	CCG	CAG	CAG	TTC	CCG	TAC	GCC	GCG	TCG	TTG	TCG	GGC	TTC	CTC	AAC
		698			707			716			725			734			743

Figure 1 (con't 2)

```
144
Pro Ser Glu Gly Trp Trp Pro Thr Leu Ile Gly Leu Ala Met Asn Asp Ser Gly
CCG TCC GAG GGC TGG TGG CCG ACG CTG ATC GGC CTG GCG ATG AAC GAC TCG GGC
      752         761         770         779         788         797
162
Gly Tyr Asn Ala Asn Ser Met Trp Gly Pro Ser Ser Asp Pro Ala Trp Lys Arg
GGT TAC AAC GCC AAC AGC ATG TGG GGT CCG TCC AGC GAC CCG GCC TGG AAG CGC
      806         815         824         833         842         851
180
Asn Asp Pro Met Val Gln Ile Pro Arg Leu Val Ala Asn Asn Thr Arg Ile Trp
AAC GAC CCA ATG GTT CAG ATT CCC CGC CTG GTC GCC AAC AAC ACC CGG ATC TGG
      860         869         878         887         896         905
198
Val Tyr Cys Gly Asn Gly Thr Pro Ser Asp Leu Gly Gly Asp Asn Ile Pro Ala
GTG TAC TGC GGT AAC GGC ACA CCC AGC GAC CTC GGC GGC GAC AAC ATA CCG GCG
      914         923         932         941         950         959
216
Lys Phe Leu Glu Gly Leu Thr Leu Arg Thr Asn Gln Thr Phe Arg Asp Thr Tyr
AAG TTC CTG GAA GGC CTC ACC CTG CGC ACC AAC CAG ACC TTC CGG GAC ACC TAC
      968         977         986         995        1004        1013
234
Ala Ala Asp Gly Gly Arg Asn Gly Val Phe Asn Phe Pro Pro Asn Gly Thr His
GCG GCC GAC GGT GGA CGC AAC GGG GTG TTT AAC TTC CCG CCC AAC GGA ACA CAC
     1022        1031        1040        1049        1058        1067
```

Figure 1 (con't 3)

```
252
Ser Trp Pro Tyr Trp Asn Glu Gln Leu Val Ala Met Lys Ala Asp Ile Gln His
TCG TGG CCC TAC TGG AAC GAG CAG CTG GTC GCC ATG AAG GCC GAT ATC CAG CAT
      1076        1085        1094        1103        1112        1121
270
Val Leu Asn Gly Ala Thr Pro Pro Ala Ala Pro Ala Ala Pro Ala Ala TER
GTG CTC AAC GGC GCG ACA CCC CCG GCC GCC CCT GCT GCG CCG GCC GCC TGAGCCA
      1130        1139        1148        1157        1166        1176

GCAAGCCAGCATCGGCAGCAGCGCAACGGCCAGCG
    1185      1194      1202      1211
```

Figure 1 (con't 4)

EP 0 499 003 A1

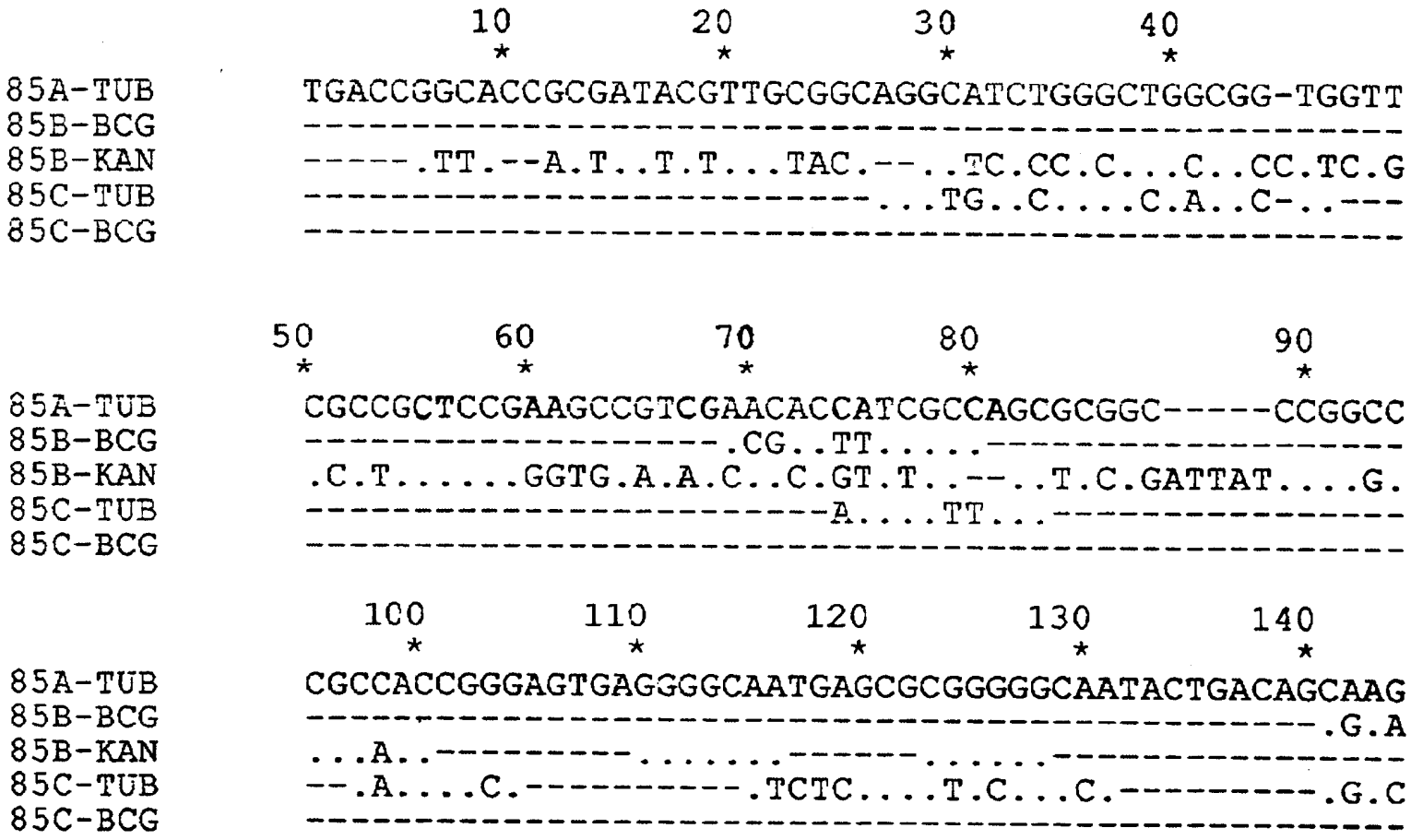

10 20 30 40
85A-TUB TGACCGGCACCGCGATACGTTGCGGCAGGCATCTGGGCTGGCGG-TGGTT
85B-BCG ----------
85B-KAN -----.TT.--A.T..T.T...TAC.--..TC.CC.C...C..CC.TC.G
85C-TUB ----------...TG..C....C.A..C-..---
85C-BCG ----------
50 60 70 80 90
85A-TUB CGCCGCTCCGAAGCCGTCGAACACCATCGCCAGCGCGGC-----CCGGCC
85B-BCG ----------.CG..TT.....----------
85B-KAN .C.T......GGTG.A.A.C..C.GT.T..--..T.C.GATTAT....G.
85C-TUB ----------A....TT..----------
85C-BCG ----------
100 110 120 130 140
85A-TUB CGCCACCGGGAGTGAGGGGCAATGAGCGCGGGGGCAATACTGACAGCAAG
85B-BCG ----------.G.A
85B-KAN ...A..---------.....------.....----------
85C-TUB --.A....C.----------.TCTC....T.C...C.---------.G.C
85C-BCG ----------

Figure 2A

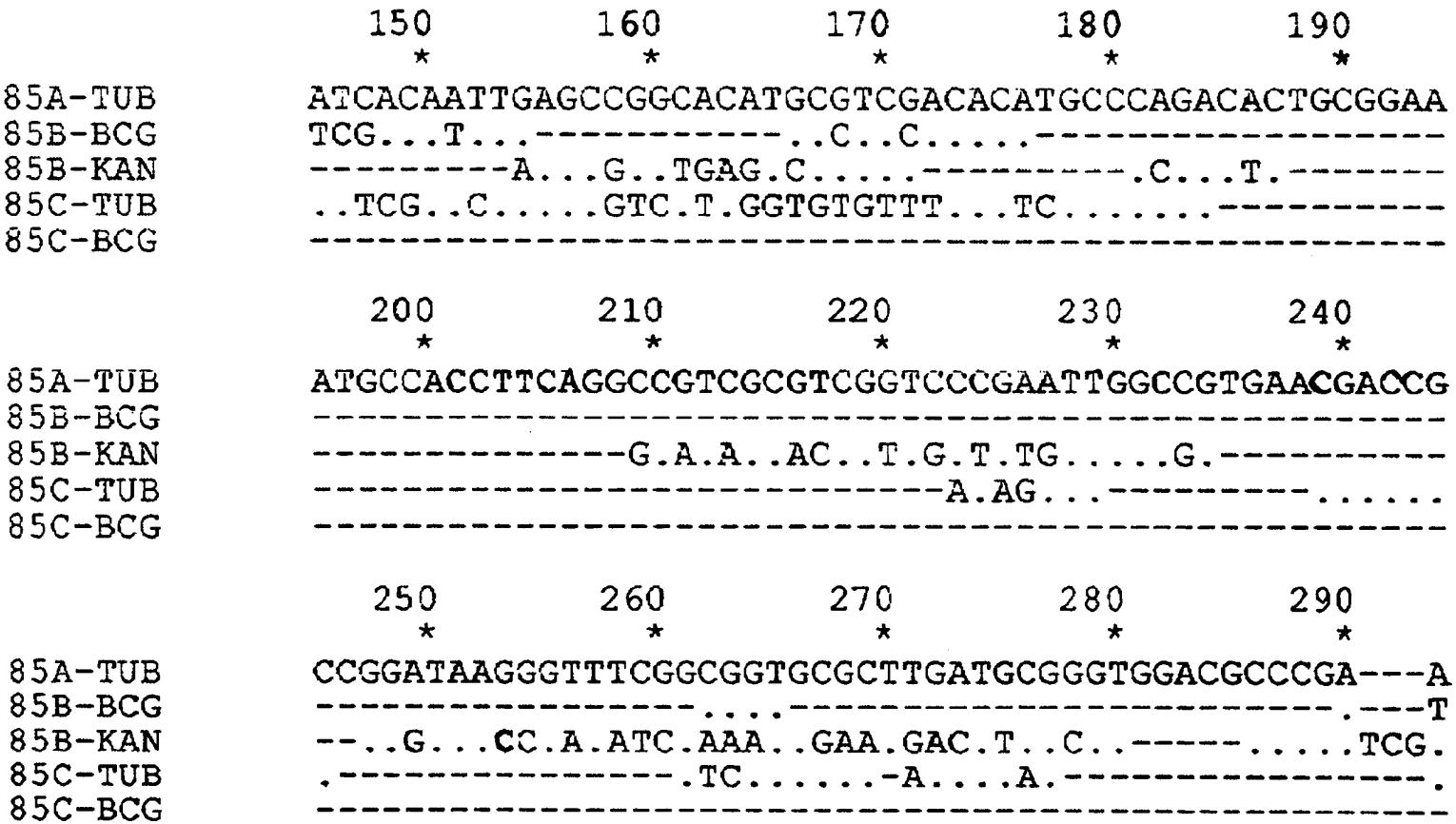


Figure 2A (con't 1)

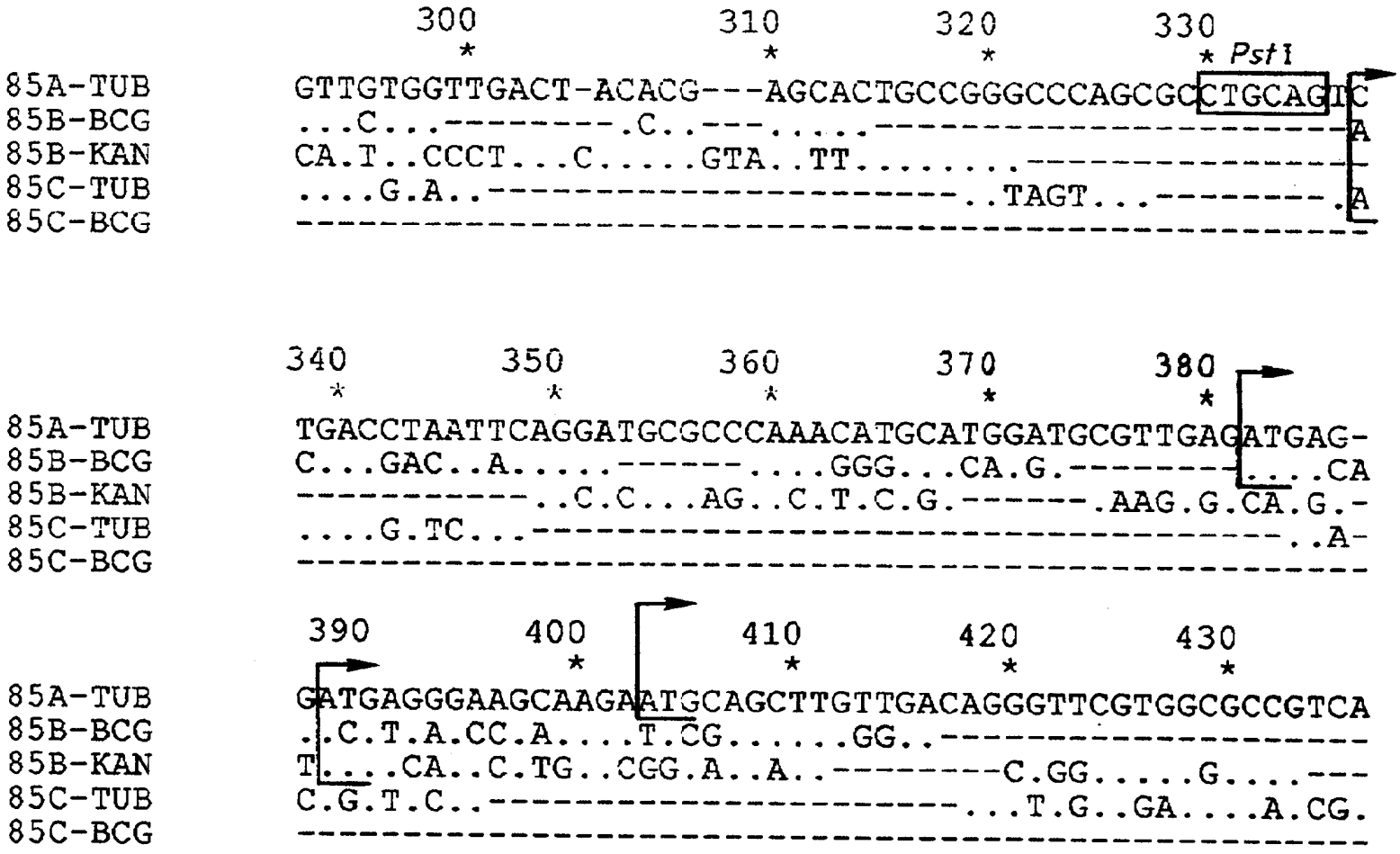


Figure 2A (con't 2)

EP 0 499 003 A1

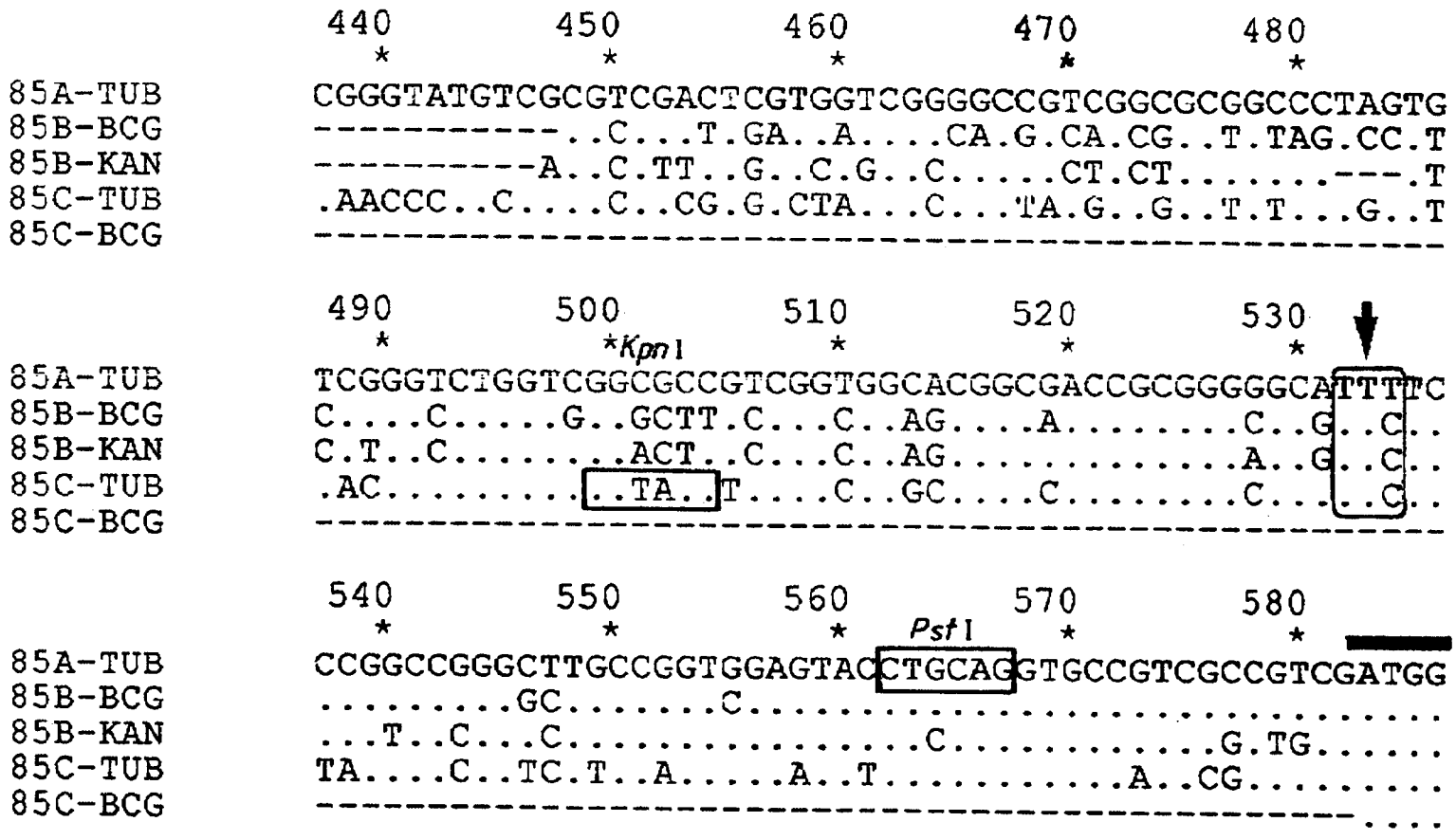


Figure 2A (con't 3)

EP 0 499 003 A1

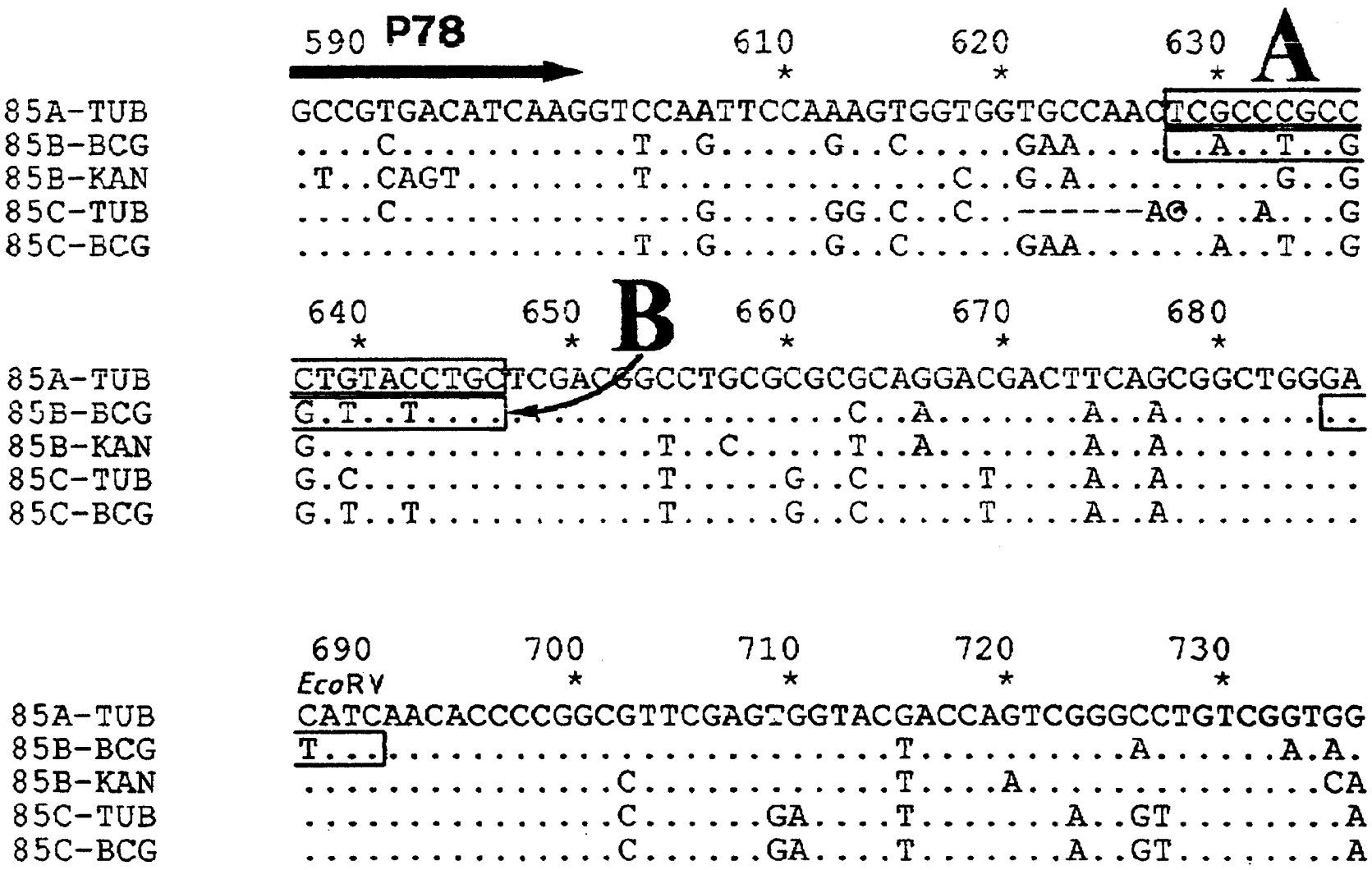


Figure 2A (con't 4)

EP 0 499 003 A1

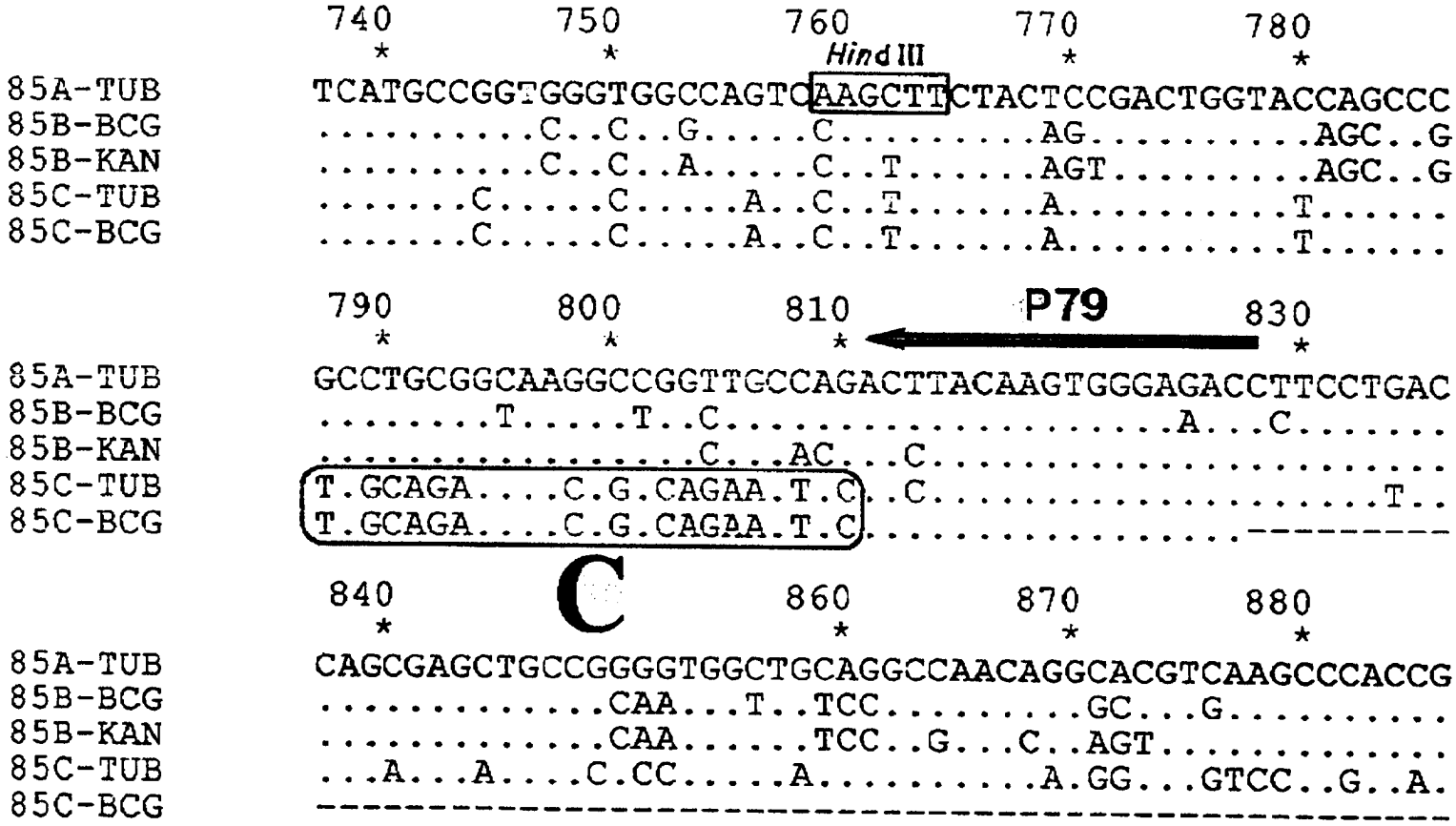


Figure 2A (con't 5)

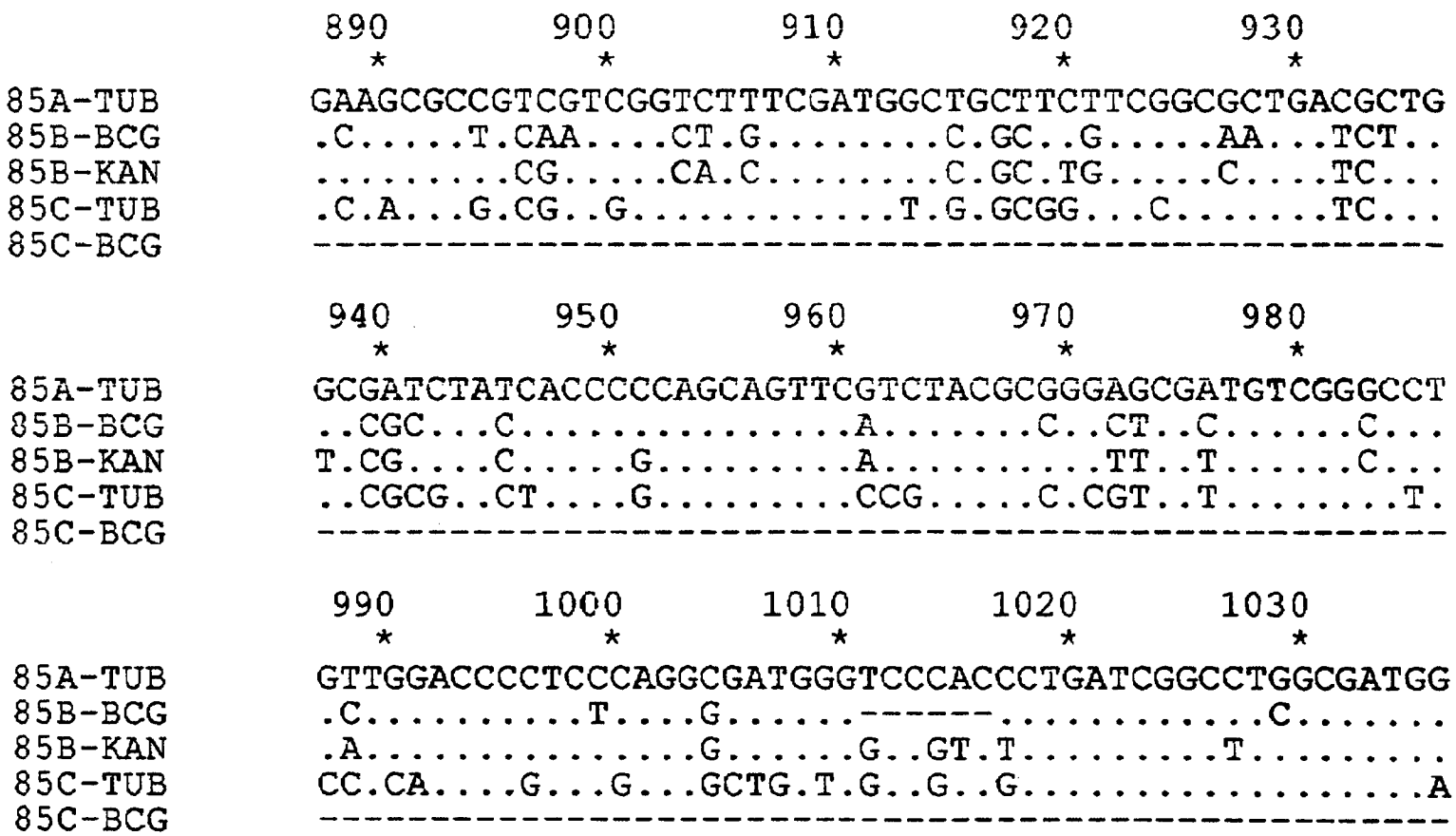


Figure 2A (con't 6)

```
                1040        1050        1060        1070        1080
                  *           *           *           *           *
85A-TUB     GTGACGCTGGCGGCTACAAGGCCTCCGACATGTGGGGCCCGAAGGAGGAC
85B-BCG     .......C.....T.........G.A...........T..CTC.AGT...
85B-KAN     .......C..T..T...........G...........A..CTC.AGT...
85C-TUB     AC...T.G.....T.....C...AA.AG.........T...TCCAGC...
85C-BCG     ------------------------------------------------

                1090        1100        1110        1120        1130
                  *           *           *           *           *
85A-TUB     CCGGCGTGGCAGCGCAACGACCCGCTGTTGAACGTCGGGAAGCTGATCGC
85B-BCG     .....A...G.............TAC.CA.C.GA..CCC......G....
85B-KAN     ..A..C........T.........TC.C..C..A.TCC.G.....G....
85C-TUB     .....C...A.............AA..G.TC.GA.TCCCCGC...G....
85C-BCG     ------------------------------------------------

                1140        1150        1160        1170        1180
                  *           *           *           *           *
85A-TUB     CAACAACACCCGCGTCTGGGTGTACTGCGGCAACGGCAAGCCGTCGGATC
85B-BCG     A...........GC.A.....T..T.....G.......CC...AAC..GT
85B-KAN     ............C.G...A.C................CC.....C..GT
85C-TUB     ............GA.................T.......CA..CAGC..C.
85C-BCG     ------------------------------------------------
```

Figure 2A (con't 7)

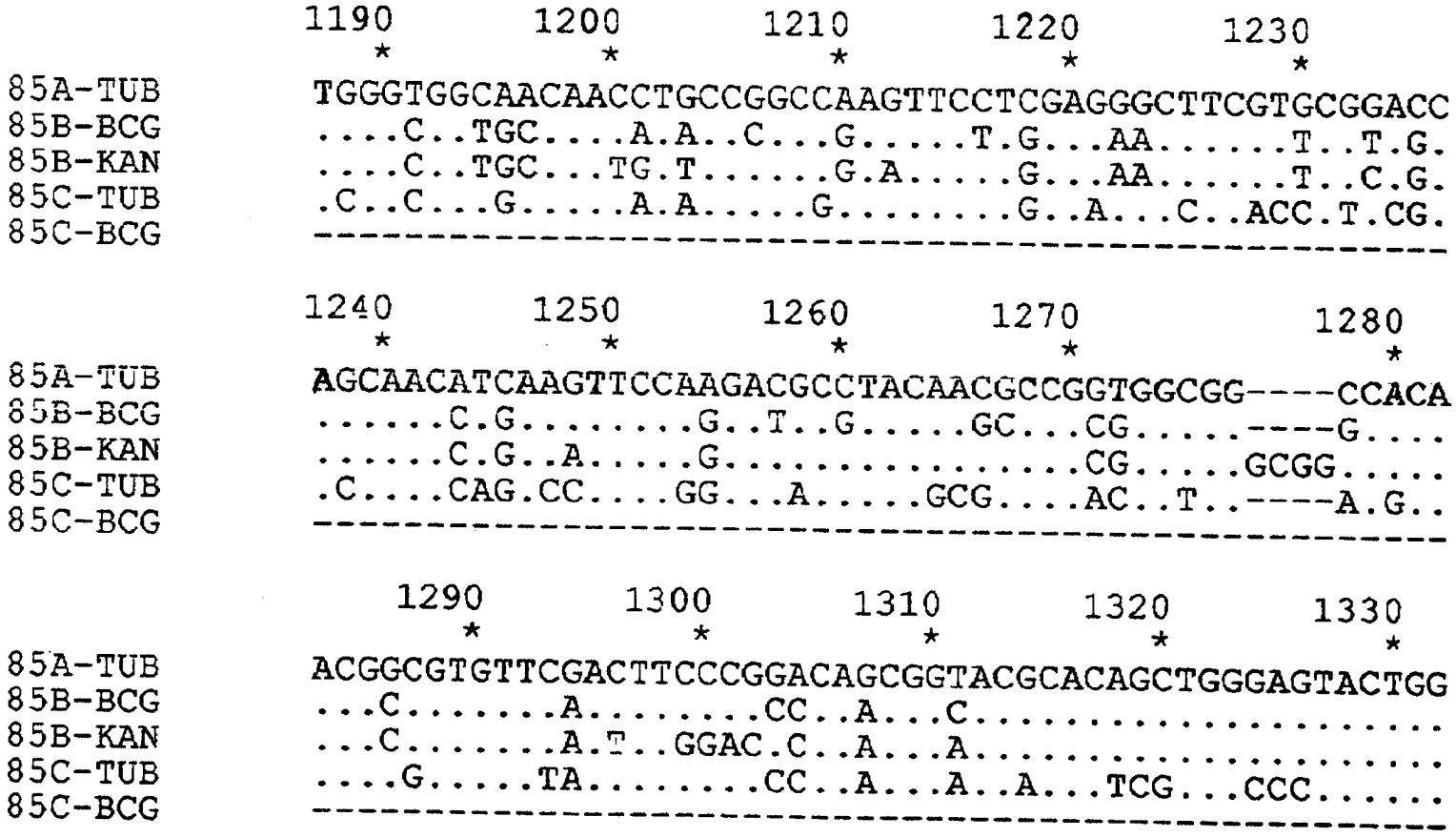


Figure 2A (con't 8)

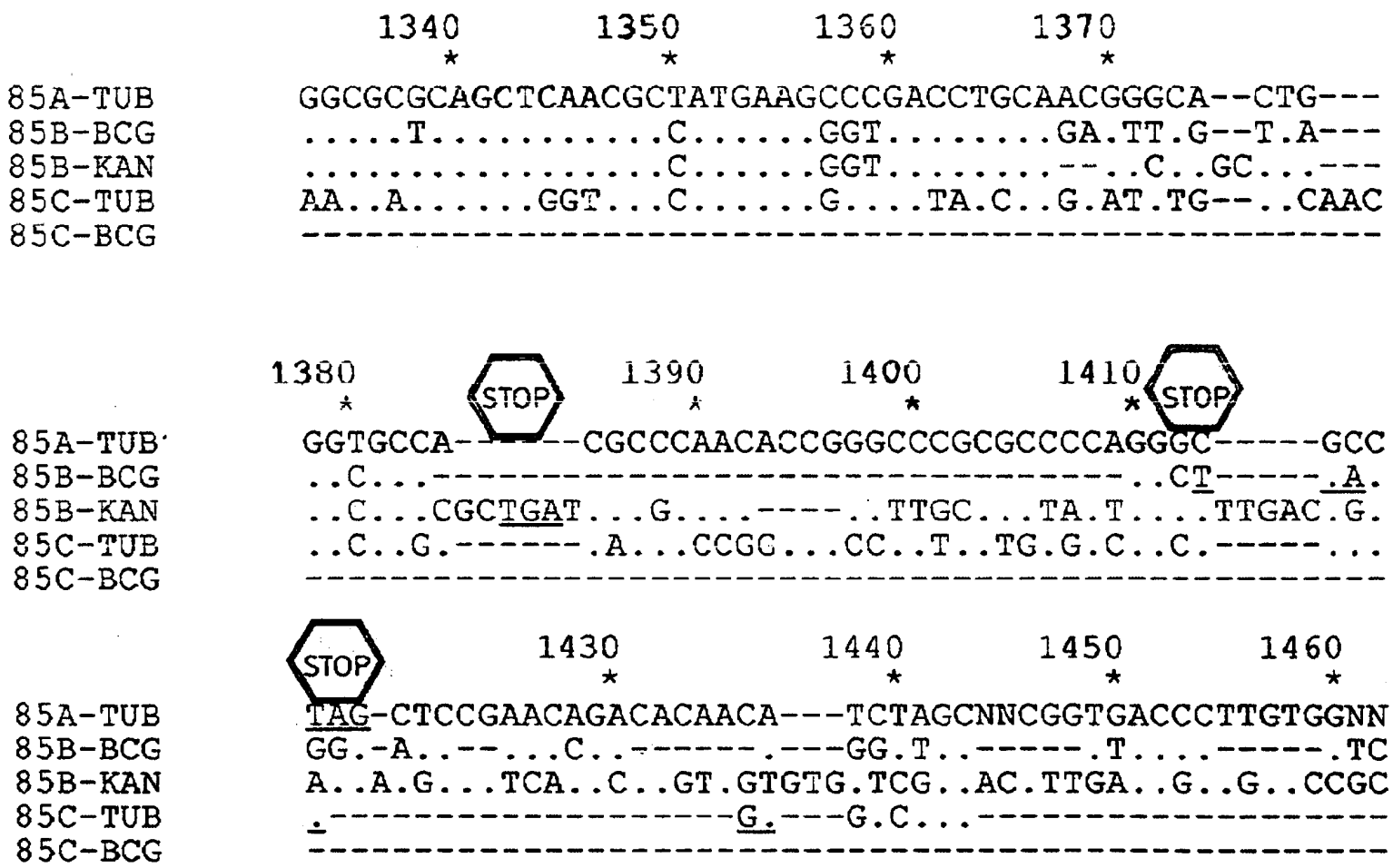


Figure 2A (con't 9)

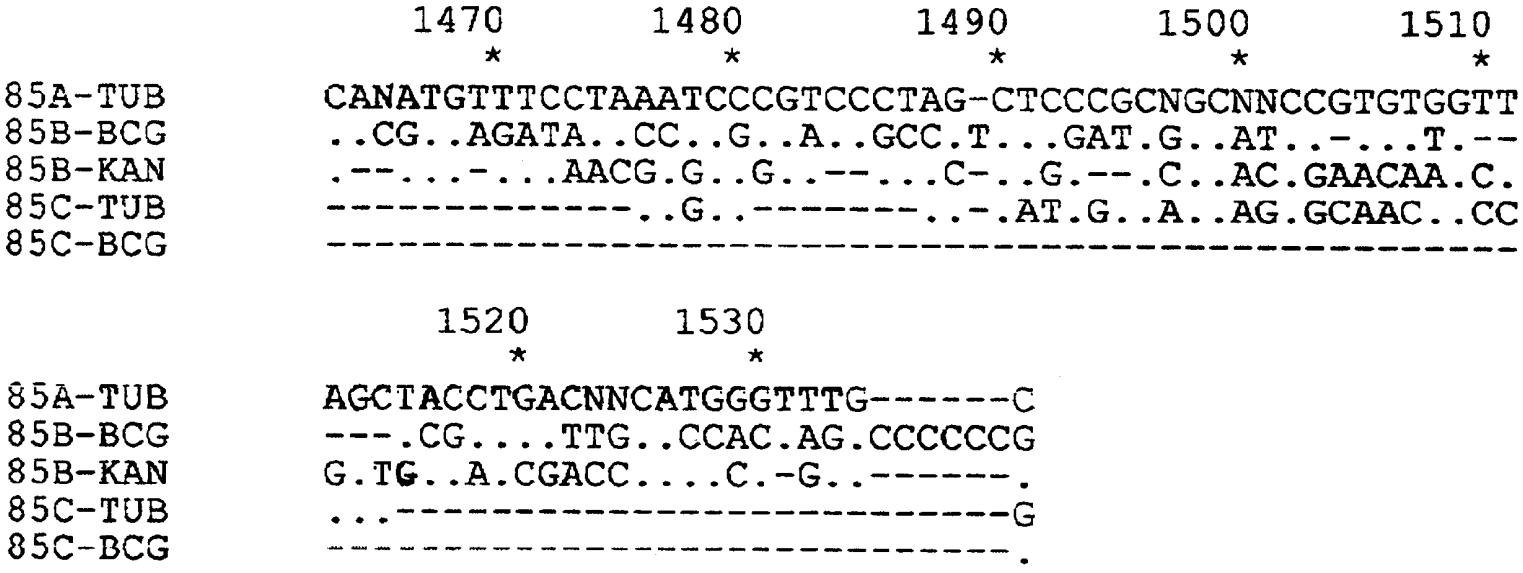


Figure 2A (con't 10)

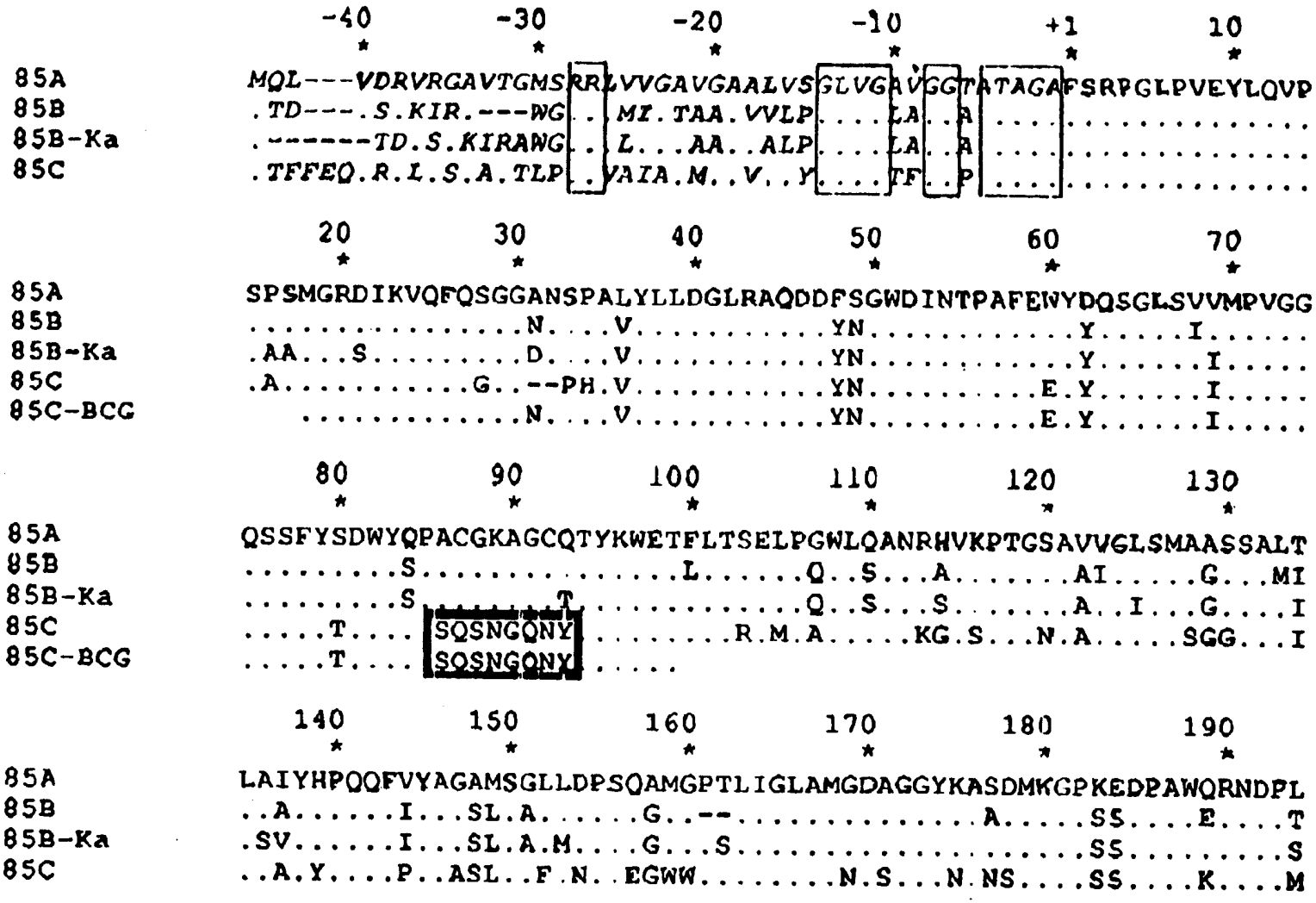


Figure 2B

EP 0 499 003 A1

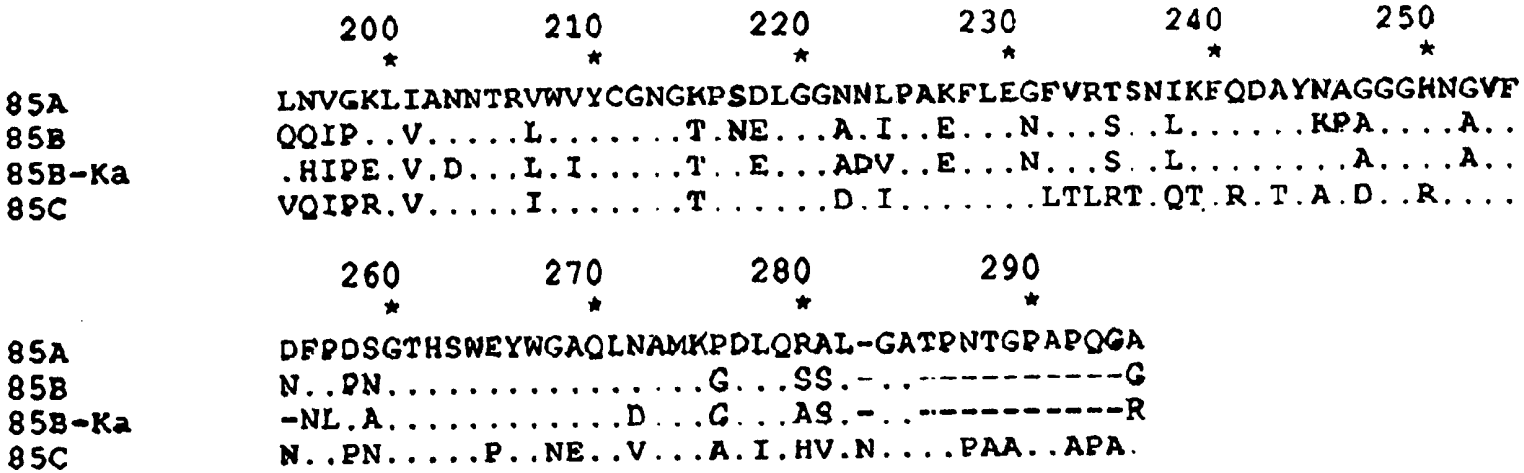


Figure 2B (con't 1)

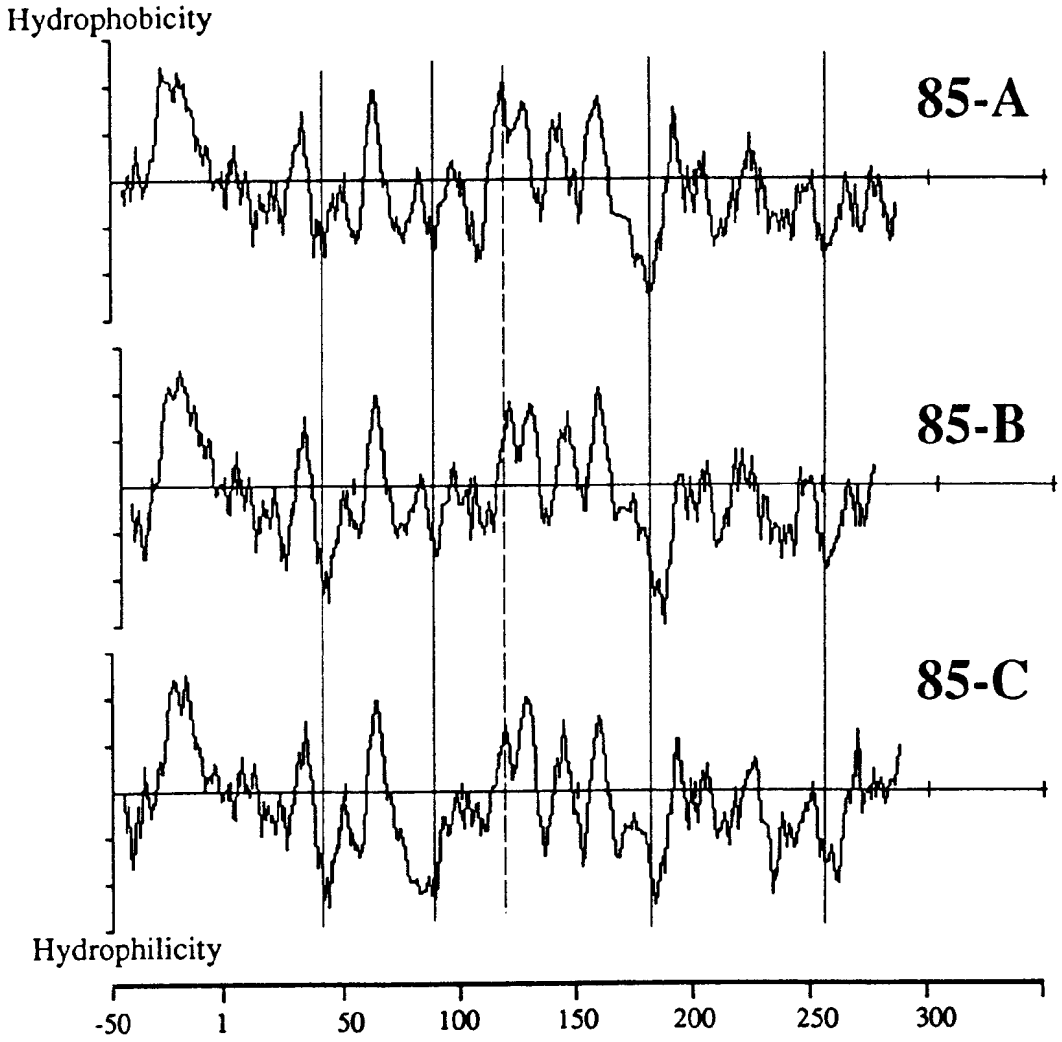

Hydrophobicity
85-A
85-B
85-C
Hydrophilicity
-50
1
50
100
150
200
250
300

Figure 3

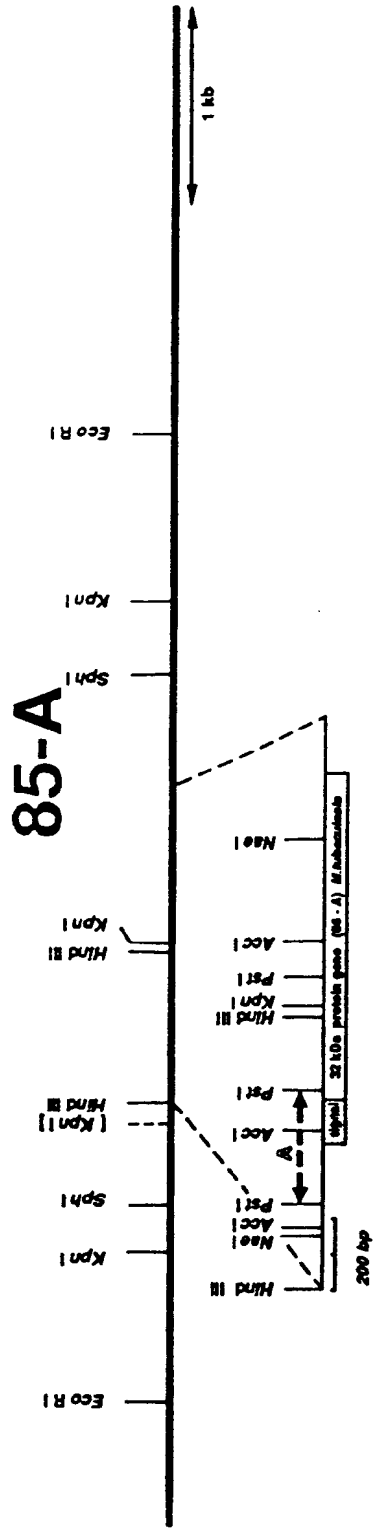
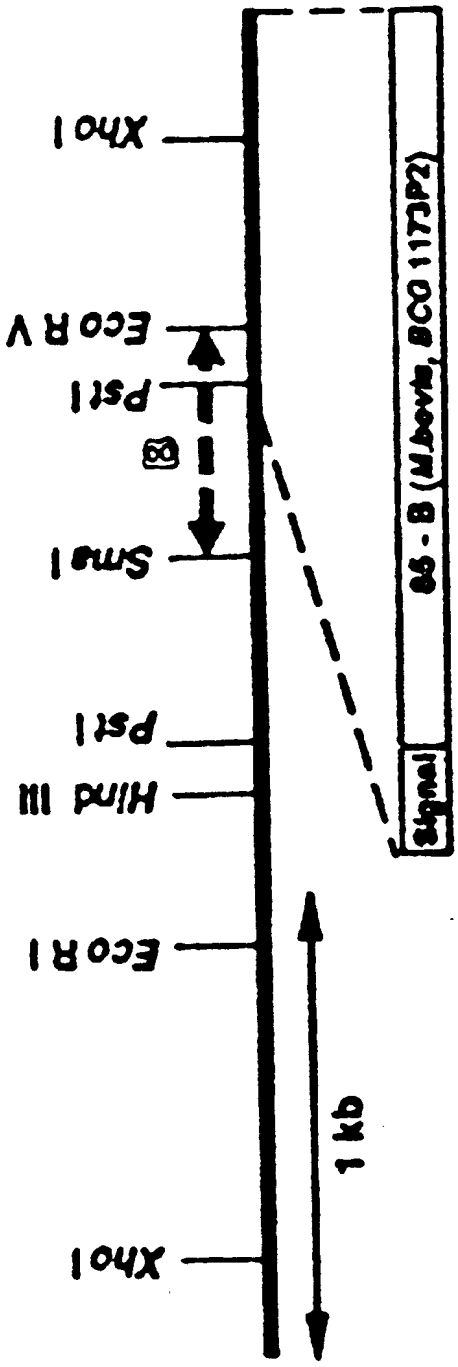

85-A
1 kb
Eco R I
Kpn I
Spn I
Kpn I
Hind III
Hind III
[Kpn I]
Spn I
Kpn I
Eco R I
Nae I
Acc I
Pst I
Kpn I
Hind III
Pst I
Acc I
A
Pst I
Nae I
Acc I
Hind III
signal
200 bp

Figure 4A



85-B

Figure 4A (con't 1)

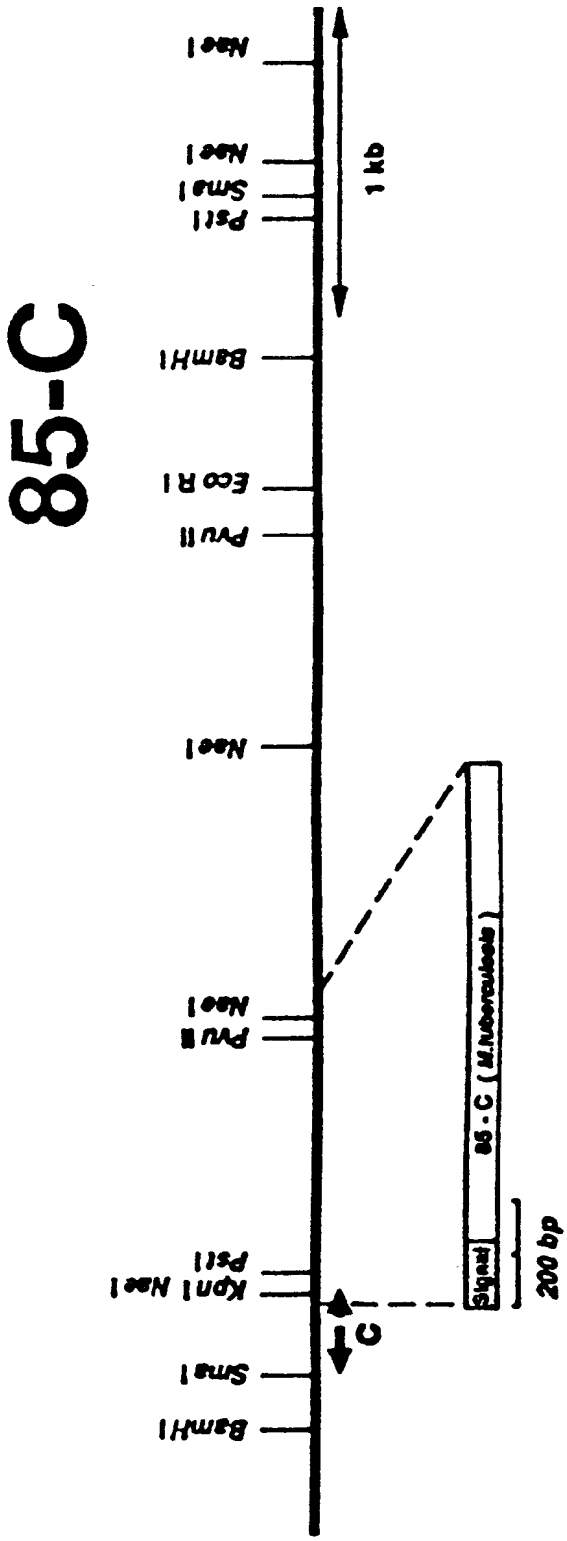


Figure 4A (con't 2)

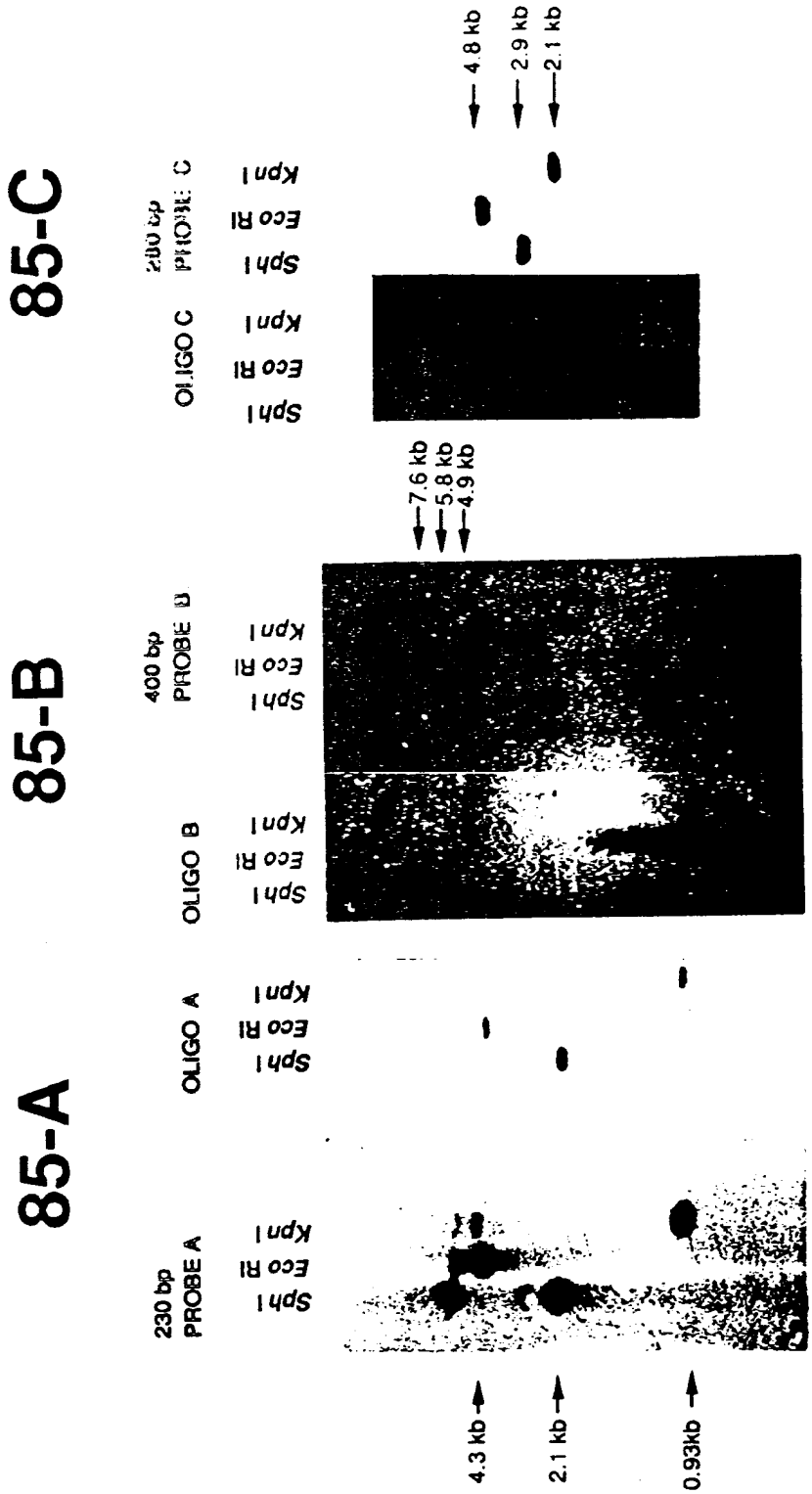

85-A
85-B
85-C
230 bp PROBE A
400 bp PROBE B
200 bp PROBE C
OLIGO A
OLIGO B
OLIGO C
Sph I
Eco RI
Kpn I
4.3 kb
2.1 kb
0.93kb
7.6 kb
5.8 kb
4.9 kb
4.8 kb
2.9 kb
2.1 kb

Figure 4B

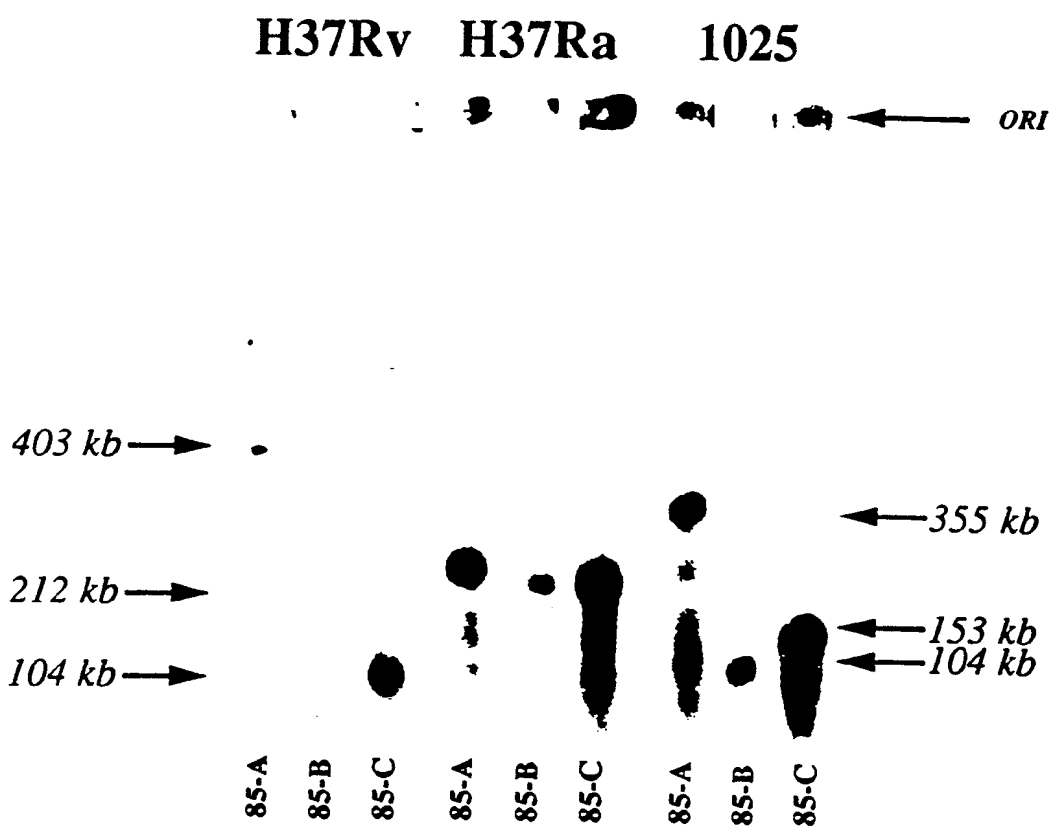

H37Rv
H37Ra
1025
ORI
403 kb
212 kb
104 kb
355 kb
153 kb
104 kb
85-A
85-B
85-C
85-A
85-B
85-C
85-A
85-B
85-C

Figure 5

European Patent Office

EUROPEAN SEARCH REPORT

Application Number

EP 91 40 0388

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int. Cl.5)
Y	INFECTION AND IMMUNITY, vol. 59, no. 1, January 1991, WASHINGTON US pages 372 - 382; S. NAGAI ET AL.: 'Isolation and Partial Characterization of Major Protein Antigens in the Culture Fluid of Mycobacterium tuberculosis ' * abstract * * table 1 * ---	1,4, 14-18	C12N15/31 C12Q1/68 G01N33/569 A61K39/04 A61K39/395 C12N15/62 C12P21/02 C12N1/21
Y	WO-A-9 010 701 (COMMONWEALTH SCIENTIFIC AND INDUSTRIAL RESEARCH ORGANISATION) September 20, 1990	1,4, 15-17	
A	* abstract * ---	19-21,24	
A	WO-A-9 012 875 (INSTITUT NATIONAL DE LA SANTE ET DE LA RECHERCHE MEDICALE (INSERM)) November 1, 1990 * page 18, line 7 - line 25 * * page 21, line 6 - line 15 * * claims 3,47 * ---	23,26	
Y	JOURNAL OF BACTERIOLOGY vol. 169, no. 3, March 1987, BALTIMORE, U.S.A pages 1080 - 1088; T. M. SHINNICK: 'The 65-Kilodalton Antigen of Mycobacterium tuberculosis ' * abstract * ---	14-18	TECHNICAL FIELDS SEARCHED (Int. Cl.5) C07K C12N A61K C12P
A	WO-A-8 701 118 (SCRIPPS CLINIC AND RESEARCH FOUNDATION) February 26, 1987 * abstract * * page 6, line 3 - line 15 * * page 8, line 26 - line 29 * * page 13, line 6 - line 22 * -----	22,25-30	

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
THE HAGUE	08 OCTOBER 1991	THIELE U.H.C.H.

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P0401)